(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 August 2004 (26.08.2004)

PCT

(10) International Publication Number
WO 2004/072279 A2

(51) International Patent Classification[7]: C12N 9/58, 9/64, 15/57, A23J 3/34

(21) International Application Number: PCT/DK2004/000091

(22) International Filing Date: 9 February 2004 (09.02.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PA 2003 00173 7 February 2003 (07.02.2003) DK
PA 2003 00426 20 March 2003 (20.03.2003) DK

(71) Applicant *(for all designated States except US)*: **NOVOZYMES A/S** [DK/DK]; Krogshoejvej 36, Bagsvaerd 2880 (DK).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PONTOPPIDAN, Katrine** [DK/DK]; Uggeloese Bygade 92a, Lynge 3540 (DK). **WU, Wenping** [CN/CN]; Room 103, Er Dan Yuan In Bldg. #3, Yi Qu, Dong Li Xiao Qu, Shang Di Zone, Haidian District (CN). **LAN, Tang** [CN/CN]; Zhichun Road 57-1-707, Hai Dian District, Beijing 100080 (CN). **LIU, Ye** [CN/CN]; #14, Zizhuyuan Lu, Haidian District, Beijing 100044 (CN). **DE MARIA, Leonardo** [IT/DK]; Kong Georgsvej 4, Frederiksberg 2000 (DK).

(74) Common Representative: **NOVOZYMES A/S**; Krogshoejvej 36, Bagsvaerd 2880 (DK).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/072279 A2

(54) Title: PROTEASES

(57) Abstract: The invention relates to a sub-class of novel proteases with distinct differences of expected structural relevance to known proteases of Metarhizium and Nocardiopsis. The sub-class encompasses proteases derived from Gliocladium, Periconia, Curvularia, and Brachysporiella. The invention also relates to DNA encoding such proteases, the expression thereof in a host cell, including animal and plant cells, as well as to the use thereof, e.g. in animal feed and in detergents. The invention furthermore relates to the use in animal feed of proteases of at least 55% identity to the protease derived from Brachysporiella. The Metarhizium protease is an The invention relates to a sub-class of novel proteases with distinct differences of expected structural relevance to known proteases of Metarhizium and Nocardiopsis. The sub-class encompasses proteases derived from Gliocladium, Periconia, Curvularia, and Brachysporiella. The invention also relates to DNA encoding such proteases, the expression thereof in a host cell, including animal and plant cells, as well as to the use thereof, e.g. in animal feed and in detergents. The invention furthermore relates to the use in animal feed of proteases of at least 55% identity to the protease derived from Brachysporiella. The Metarhizium protease is an example of such proteases.

Proteases

Field of the Invention

The present invention relates to a sub-class of novel serine proteases of peptidase families S2A and/or S1E that differ from related known proteases. A multiple alignment of these proteases, shown in Figure 1, reveals distinct differences of expected structural relevance. In particular, using the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6) as shown in the upper row of this alignment, the amino acid sequences of the proteases of the invention do not comprise any of the following combinations of amino acids: (K114 + P121 + R130 + S162 + R163 + R174 + Q177), and (R114 + P121 + R130 + R162 + T163 + P174 + N177). The invention also relates to DNA encoding these proteases, their recombinant production, and their use in animal feed and detergents.

Furthermore, the invention relates to the use in animal feed of proteases homologous to a *Brachysporiella* protease (examples of which derives from *Brachysporiella* (*Ascotaiwania*) and *Metarhizium*).

Background of the Invention

The cloning and expression of a protease derived from *Metarhizium anisopliae* is disclosed by Steven E. Screen and Raymond J. St. Leger in The Journal of Biological Chemistry, Vol. 275, No. 9, 2000, pp 6689-6694. The nucleotide sequence, chy1, thereof is shown in the sequence listing as SEQ ID NO: 3, and the deduced amino acid sequence, CHY1, as SEQ ID NO: 4 (TREMBL:Q9Y843).

Proteases derived from *Nocardiopsis sp.* NRRL 18262 and *Nocardiopsis dassonvillei* NRRL 18133 are described in WO 88/03947. The DNA and amino acid sequences of the protease derived from *Nocardiopsis sp.* NRRL 18262 are shown in DK patent application no. 1996 00013. WO 01/58276 describes the use in animal feed of acid-stable proteases related to the protease derived from *Nocardiopsis sp.* NRRL no. 18262. JP 2255081 A describes a protease purified from *Nocardiopsis sp.* FERM P-1-508. GDR patent no. DD 2,004,328 discloses a protease derived from *Nocardiopsis dassonvillei* ZIMET 43647.

It is an object of the present invention to provide alternative proteases for various industrial uses, for example for use in animal feed and/or detergents.

Summary of the Invention

The present invention relates to an isolated polypeptide having protease activity, for short a protease, the protease belonging to peptidase family S2A and/or peptidase family S1E, and comprising an amino acid sequence which, when aligned according to Fig. 1, does

not comprise any of the following:

a) (K114+P121+R130+S162+R163+R174+Q177); and not

b) (R114+P121+R130+R162+T163+P174+N177).

wherein the numbering of each amino acid residue corresponds to the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6).

The invention also relates to isolated nucleic acid sequences encoding the polypeptides and to nucleic acid constructs, vectors, and host cells comprising the nucleic acid sequences as well as methods for producing and using the polypeptides.

Brief Description of the Figures

Figure 1 is a multiple alignment made using the program CLUSTAL W (1.82) of proteases derived from *Gliocladium sp.* CBS 114001 (amino acids 1-186 of SEQ ID NO: 6), *Periconia sp.* CBS 114000 (amino acids 1-65 of SEQ ID NO: 12), *Curvularia lunata* CBS 114003 (amino acids 1-185 of SEQ ID NO: 10), *Periconia sp.* CBS 114002 (amino acids 1-186 of SEQ ID NO: 8), *Brachysporiella gayana* CGMCC 0865 (amino acids 1-186 of SEQ ID NO: 2), *Metarhizium anisopliae* TREMBL:Q9Y843 (amino acids 1-188 of SEQ ID NO: 4), and *Nocardiopsis sp.* NRRL 18262 (amino acids 1-188 of SEQ ID NO: 26). The signal peptide parts and pro-peptide parts of these proteases were included when preparing the alignments but have been deleted from Fig. 1.

Detailed Description of the Invention

A description of serine proteases of peptidase families S2A and S1E is included in the section headed "polypeptides having protease activity."

Amino Acid Residue Numbering

In the present context, for the purposes of identifying corresponding amino acid residues in various proteases, reference is had to the numbering of the amino acid residues in the mature part (amino acids 1-186) of SEQ ID NO: 6, the *Gliocladium* protease, starting with V1 and ending with A186. Fig. 1 shows the numbers of the last amino acid residues of each of the top rows of the alignment. For example, the number "41" means that the last amino acid residue of this first top row of the alignment, "S", is amino acid residue number 41 in the mature *Gliocladium* protease amino acid sequence. As another example, the number "160" means that the last amino acid residue of this third top row of the alignment of Fig. 1, which by the way is also an "S," is residue number 160 in the mature *Gliocladium* protease amino acid sequence. This numbering is identical to the numbering of SEQ ID NO: 6, however not necessarily identical to the numbering of SEQ ID NOs: 12, 10, 8, 2, 4, and 26, the reason being that for the purposes of identifying corresponding amino acid residues

in various proteases, a uniform numbering is used based on the *Gliocladium* protease. The procedure for assigning a uniform numbering is further described below.

For each of the amino acid residues of the *Gliocladium* protease, a "corresponding" residue can be identified in each of the other six proteases shown in Fig. 1, because "corresponding" residues are simply those that are placed one above the other, or on top of each other, in the alignment of Fig. 1. For example, the first amino acid residue, V, of the *Gliocladium* protease (V1 of SEQ ID NO: 6) corresponds to V1 of SEQ ID NOs: 12, 10, and 8; to I1 of SEQ ID NO: 2; to V3 of SEQ ID NO: 4; and to I3 of SEQ ID NO: 26. In the present context, however, for all purposes except for Sequence Listing purposes, these residues are all assigned the same number, because they qualify as "corresponding residues," and the number assigned is that of the corresponding residue in the *Gliocladium* protease, viz. residue number 1. Accordingly, V1 of the *Gliocladium* protease corresponds to V1, V1, V1, I1, V1, and I1, of the *Periconia* CBS 114000, *Curvularia*, *Periconia* CBS 114002, *Brachysporiella*, *Metarhizium*, and *Nocardiopsis* proteases, respectively. When nothing else is stated, this numbering is used hereinafter.

The multiple alignment of Fig. 1, in certain rows, at certain positions, includes gaps, which can be considered as deletions of amino acid residues. In the present context, the gaps, or the deleted amino acid residues, are numbered by assigning to each gap lower case letters in alphabetical order, viz. a, b, c, d, ---- t, u, v, x, y, z. Should more than 25 of such designations be needed, the numbering would continue with aa, bb, cc etc. For example, the gap between G24 and F25 of the *Gliocladium* protease corresponds to a deletion of five amino acid residues in positions 24a, 24b, 24c, 24d, and 24e. Accordingly, the corresponding residues in the *Nocardiopsis* protease (SEQ ID NO: 26) shown in the last row of Fig. 1 are designated A24a, G24b, Q24c, P24d, and G24e. By analogy, the successive amino acid residues PGND in the *Periconia* CBS 114000 sequence corresponding to PGSAD in positions 55-59 of the *Gliocladium* protease are numbered as follows: P55, G56, N57, and D59; position 58 in this protease being equivalent to a deletion of an amino acid.

Two of the proteases included in the alignment of Fig. 1 comprise N-terminal extensions as compared to SEQ ID NO: 6, viz. the *Metarhizium* and the *Nocardiopsis* proteases of the last two rows. The amino acids of such extensions are numbered as is usual in the art, -1, -2, -3 and so forth ("-" meaning "minus"). For example, the first three amino acid residues of the mature *Nocardiopsis* protease are referred to as A-2, D-1, and I1, respectively.

Another protease having an amino acid sequence of SEQ ID NO: X (including signal peptide, pro-peptide, and mature peptide parts) may be added to the alignment of Fig. 1 as follows:

The percentage of identity of SEQ ID NO: X to each of the seven proteases of Fig. 1 (each of SEQ ID NOs: 6, 12, 10, 8, 2, 4, and 25) is determined using the "Align" program as described below. Seven pairwise alignments are thereby produced. The sequence with the highest degree of identity to SEQ ID NO: X is selected as a model protease. If there are more candidate model proteases, you select the one which is listed first in the Fig. 1 alignment. If, for example, SEQ ID NO: X would have the following percentage of identities to SEQ ID NOs: 6, 12, 10, 8, 2, 4, and 25: 45%, 55%, 50%, 65%, 65%, 45%, and 40%, respectively, then SEQ ID NO: 8 should be selected as the model protease. As a next step, using the pairwise alignment of SEQ ID NO: X to SEQ ID NO: 8, SEQ ID NO: X is pasted (or it could simply be written) onto the alignment of Fig. 1 as the bottom row, ensuring that corresponding amino acid residues (here "corresponding" refers to the pairwise alignment of SEQ ID NO: X and SEQ ID NO: 8) are placed above each other.

While the Fig. 1 alignment remains unaffected by this procedure of adding a new sequence to it, the described procedure may give rise to gaps in SEQ ID NO: X; "loops" in SEQ ID NO: X, and/or N- or C-terminal extensions of SEQ ID NO: X, as compared to the *Gliocladium* protease of Fig. 1.

As regards the numbering of such positions with a view to identifying corresponding amino acid residues in SEQ ID NO: X, the gaps and the N-terminal extensions are dealt with as described above. C-terminal extensions, if any, are numbered as is usual in the art, i.e. continuing from residue no. 186 with 187, 188, 189 and so forth. As regards the loops, if any, this corresponds to SEQ ID NO: X having "excess" amino acid residues, which the alignment of Fig. 1 does not make room for. Typographically, such excess residues are transferred onto a next row, but they are of course considered to be included in the multiple alignment, and are numbered by analogy to what is described above for the numbering of gaps (using the denotations a, b, c etc.). For example, the pairwise alignment of SEQ ID NO: X and SEQ ID NO: 8 could include the following part:

```
(part of SEQ ID NO: 8)      CNS----GGTTFFQPVNE
(part of SEQ ID NO: X)      CRTAKSAGGQTYFQEVTE,
```

The amino acid residues CNSGGTTFFQPVNE of the *Periconia* CBS 114002 protease (SEQ ID NO: 8) have alignment numbers 161 to 174 and these turn up to correspond, in the pairwise alignment, to the following amino acid residues of SEQ ID NO: X: CRTAKSAGGQTYFQEVTE. For the present purposes, the amino acid residues of SEQ ID NO: X would have to be numbered as follows: C161, R162, T163, A163a, K163b, S163c, A163d, G164, G165, Q166, T167, Y168, F169, Q170, E171, V172, T173, and E174.

In the alignment of Fig. 1 this may be written as follows:

```
(part of SEQ ID NO: 8)      CNSGGTTFFQPVNE
(part of SEQ ID NO: X)      CRTGGQTYFQEVTE
                              AKSA 163d
```

In a particular alternative embodiment, the multiple alignment of Fig. 1 is produced using the mature part of SEQ ID NOs: 6, 12, 10, 8, 2, 4, and 26. In another particular embodiment, the pairwise alignments referred to above are produced using the mature part of SEQ ID NO:X, and mature parts of SEQ ID NOs: 6, 12, 10, 8, 2, 4, and 26, respectively.

The following are examples of the denotation used herein to characterize proteases:

Expressions like the following:

"A protease comprising an amino acid sequence which, when aligned according to Fig. 1, comprises S61, wherein the numbering of the amino acid residue corresponds to the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6)"

mean, that the amino acid sequence of the protease in question, when added to the multiple alignment of Fig. 1 as described above, in position number 61 (which refers to the alignment number deduced as described above) has an S (Ser, Serine).

Expressions like "CDEFGHIKLMNPQRSTVWY61" mean, that the amino acid residue in position number 61 can be either of C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, or Y.

Expressions like "at least one of (L114 or Y114); (S121 or D121); (Q130 or M130); (N162 or T162 or S162); (S163 or R163); E174; and/or (S177 or E177)," are used about a protease that fulfills either of (at least one of) the criteria separated by semicolons (;). This is the case, for example, for a protease having Y114, or S121, or M130, or E174, but also for a protease having L114 and D121.

Expressions like "((L114 or Y114) + (S121 or D121) + (Q130 or M130) + (N162 or T162 or S162) + (S163 or R163) + E174 + (S177 or E177))," are used about a protease that fulfills all of the criteria separated by plusses (+). This is the case, for example, for a protease having Y114 and S121 and M130 and T162 and R163 and E174 and E177.

Expressions like "(H30 + D59 + S140)," is used about a protease comprising H30, D59 and S140.

Structural Considerations

We identified seven positions in the alignment of Fig. 1 of a potential high structural relevance, viz. positions 114, 121, 130, 162, 163, 174, and 177. The differences observed in these positions between the known *Metarhizium* and *Nocardiopsis* proteases and the proteases of the invention are striking and may give rise to interesting characteristic features of the enzymes of the invention. Also in other positions, characteristic differences are observed, viz. in positions 61, 63, 77, and 111.

The catalytic triad of the proteases of the invention is constituted by amino acids H30, D59, and S140.

Polypeptides having Protease Activity

Polypeptides having protease activity, or proteases, are sometimes also designated peptidases, proteinases, peptide hydrolases, or proteolytic enzymes. Proteases may be of the exo-type that hydrolyses peptides starting at either end thereof, or of the endo-type that act internally in polypeptide chains (endopeptidases). Endopeptidases show activity on N- and C-terminally blocked peptide substrates that are relevant for the specificity of the protease in question.

The term "protease" is defined herein as an enzyme that hydrolyses peptide bonds. It includes any enzyme belonging to the EC 3.4 enzyme group (including each of the thirteen subclasses thereof). The EC number refers to Enzyme Nomenclature 1992 from NC-IUBMB, Academic Press, San Diego, California, including supplements 1-5 published in Eur. J. Biochem. 1994, 223, 1-5; Eur. J. Biochem. 1995, 232, 1-6; Eur. J. Biochem. 1996, 237, 1-5; Eur. J. Biochem. 1997, 250, 1-6; and Eur. J. Biochem. 1999, 264, 610-650; respectively. The nomenclature is regularly supplemented and updated; see e.g. the World Wide Web (WWW) at http://www.chem.qmw.ac.uk/iubmb/enzyme/index.html).

Proteases are classified on the basis of their catalytic mechanism into the following groups: Serine proteases (S), Cysteine proteases (C), Aspartic proteases (A), Metalloproteases (M), and Unknown, or as yet unclassified, proteases (U), see Handbook of Proteolytic Enzymes, A.J.Barrett, N.D.Rawlings, J.F.Woessner (eds), Academic Press (1998), in particular the general introduction part.

The proteases of the invention are selected from the group consisting of:

(a) Serine proteases of peptidase family S2A according to the above Handbook; and

(b) Serine proteases of peptidase family S1E as described in Biochem.J. 290:205-218 (1993) and in MEROPS protease database, release 6.20, March 24, 2003, (www.merops.ac.uk). The database is described in Rawlings, N.D., O'Brien, E. A. & Barrett, A.J. (2002) MEROPS: the protease database. Nucleic Acids Res. 30, 343-346.

Peptidase family S2A represents the traditional way of classifying proteases. Nowadays, proteases traditionally classified as S2A proteases are often classified according to the MEROPS classification in peptidase family S1E.

In a particular embodiment, the protease is of peptidase family S2A. In another particular embodiment, the protease is of peptidase family S1E.

In alternative embodiments, the proteases of the invention are selected from the group consisting of:

(c) proteases belonging to the EC 3.4.-.- enzyme group; and

(d) Serine proteases belonging to the S group of the above Handbook;

For determining whether a given protease is a Serine protease, a family S2A protease, and/or a family S1E protease, reference is made to the above references and the principles indicated therein. Such determination can be carried out for all types of proteases, be it naturally occurring or wild-type proteases; or genetically engineered or synthetic proteases.

Protease activity can be measured using any assay, in which a substrate is employed, that includes peptide bonds relevant for the specificity of the protease in question. Assay-pH and assay-temperature are likewise to be adapted to the protease in question. Examples of assay-pH-values are pH 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12. Examples of assay-temperatures are 20, 25, 30, 35, 37, 40, 45, 50, 55, 60, 65, 70 or 80°C.

Examples of protease substrates are casein, such as Azurine-Crosslinked Casein (AZCL-casein). Two protease assays are described in Example 1 herein, either of which can be used to determine protease activity. For the purposes of this invention, the so-called pNA Assay is a preferred assay.

There are no limitations on the origin of the protease of the invention and/or for use according to the invention. Thus, the term protease includes not only natural or wild-type proteases obtained from microorganisms of any genus, but also any mutants, variants, fragments etc. thereof exhibiting protease activity, as well as synthetic proteases, such as shuffled proteases, and consensus proteases. Such genetically engineered proteases can be prepared as is generally known in the art, eg by Site-directed Mutagenesis, by PCR (using a PCR fragment containing the desired mutation as one of the primers in the PCR reactions), by shuffling, or by Random Mutagenesis. The preparation of consensus proteins is described in eg EP 897985. The term "obtained from" as used herein in connection with a given source shall mean that the polypeptide encoded by the nucleic acid sequence is produced by the source or by a cell in which the nucleic acid sequence from the source is present. In a preferred embodiment, the polypeptide is secreted extracellularly.

In a specific embodiment, the protease is a low-allergenic variant, designed to invoke a reduced immunological response when exposed to animals, including man. The term immunological response is to be understood as any reaction by the immune system of an animal exposed to the protease. One type of immunological response is an allergic response leading to increased levels of IgE in the exposed animal. Low-allergenic variants may be prepared using techniques known in the art. For example the protease may be conjugated with polymer moieties shielding portions or epitopes of the protease involved in an immunological response. Conjugation with polymers may involve *in vitro* chemical coupling of polymer to the protease, e.g. as described in WO 96/17929, WO 98/30682, WO

98/35026, and/or WO 99/00489. Conjugation may in addition or alternatively thereto involve in vivo coupling of polymers to the protease. Such conjugation may be achieved by genetic engineering of the nucleotide sequence encoding the protease, inserting consensus sequences encoding additional glycosylation sites in the protease and expressing the protease in a host capable of glycosylating the protease, see e.g. WO 00/26354. Another way of providing low-allergenic variants is genetic engineering of the nucleotide sequence encoding the protease so as to cause the protease to self-oligomerize, effecting that protease monomers may shield the epitopes of other protease monomers and thereby lowering the antigenicity of the oligomers. Such products and their preparation is described e.g. in WO 96/16177. Epitopes involved in an immunological response may be identified by various methods such as the phage display method described in WO 00/26230 and WO 01/83559, or the random approach described in EP 561907. Once an epitope has been identified, its amino acid sequence may be altered to produce altered immunological properties of the protease by known gene manipulation techniques such as site directed mutagenesis (see e.g. WO 00/26230, WO 00/26354 and/or WO 00/22103) and/or conjugation of a polymer may be done in sufficient proximity to the epitope for the polymer to shield the epitope.

In a particular embodiment, the polypeptide of the invention comprises an amino acid sequence which has a degree of identity to the mature peptide parts of any one of SEQ ID NOs: 2, 6, 8, 10, and/or 12 of at least about 40%, and which have protease activity (hereinafter "homologous polypeptides"). In further particular embodiments, the degree of identity is at least about 42%, 44%, 46%, 48%, 50%, 52%, 54%, 56%, 58%, 60%, 62%, 63%, 64%, 66%, 68%, 70%, 72%, 75%, 77%, 80%, 82%, 85%, 87%, 90%, 92%, 95%, or at least about 97%. In another alternative embodiment, any of the above degrees of identity is relative to amino acids -170 to 186, or -189 to 186 of SEQ ID NO: 2 or the corresponding parts of any one of SEQ ID NOs: 2, 6, 8, 10, and/or 12. In particular embodiments, the polypeptides of the invention i) have; or ii) consist of an amino acid sequence with any of the degrees of identity as mentioned above.

For the purposes of the present invention, the degree of identity between two amino acid sequences, as well as the degree of identity between two nucleotide sequences, is determined by the program "align" which is a Needleman-Wunsch alignment (i.e. a global alignment). The program is used for alignment of polypeptide, as well as nucleotide sequences. The default scoring matrix BLOSUM50 is used for polypeptide alignments, and the default identity matrix is used for nucleotide alignments. The penalty for the first residue of a gap is -12 for polypeptides and -16 for nucleotides. The penalties for further residues of a gap are -2 for polypeptides, and -4 for nucleotide.

"Align" is part of the FASTA package version v20u6 (see W. R. Pearson and D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444-2448, and W. R. Pearson (1990) "Rapid and Sensitive Sequence Comparison with FASTP and FASTA," Methods in Enzymology 183:63-98). FASTA protein alignments use the Smith-Waterman algorithm with no limitation on gap size (see "Smith-Waterman algorithm", T. F. Smith and M. S. Waterman (1981) J. Mol. Biol. 147:195-197).

In a particular embodiment, the mature peptide parts, or predicted or expected mature peptide parts, of the two amino acid sequences are used for the alignment.

The present invention also relates to the animal feed use of a protease having at least 55% identity to amino acids 1-186 of SEQ ID NO: 2. In particular embodiments of this use, the degree of identity is at least about 60%, 62%, 63%, 64%, 65%, 67%, 70%, 72%, 75%, 77%, 80%, 82%, 85%, 87%, 90%, 92%, 95%, or at least about 97%. In one alternative embodiment of this use, the degree of identity is at least about 50%, 51%, 52%, 53%, or at least about 54%. The degree of identity is determined as described above.

The percentage of identity of the mature peptide part of SEQ ID NO: 4 to the mature peptide part of SEQ ID NO: 2 is 61.2%. The percentage of identity of the mature peptide part of SEQ ID NO: 2 to the mature peptide part of a protease derived from *Nocardiopsis sp.* NRRL 18262 (having SEQ ID NO: 1 of WO 01/58276) is 47.4%.

The degree of identity between two amino acid sequences may also be determined by the Clustal method (Higgins, 1989, CABIOS 5: 151-153) using the LASERGENE™ MEGALIGN™ software (DNASTAR, Inc., Madison, WI) with an identity table and the following multiple alignment parameters: Gap penalty of 10, and gap length penalty of 10. Pairwise alignment parameters are Ktuple=1, gap penalty=3, windows=5, and diagonals=5. The degree of identity between two nucleotide sequences may be determined using the same algorithm and software package as described above with the following settings: Gap penalty of 10, and gap length penalty of 10. Pairwise alignment parameters are Ktuple=3, gap penalty=3 and windows=20.

In a particular embodiment, the homologous polypeptides have an amino acid sequence that differs by ten, or by nine, or by eight, or by seven, or by six, or by five amino acids. In another particular embodiment, the homologous polypeptides differ by four, or by three, or by two amino acids, or by one amino acid from amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2 or the corresponding parts of SEQ ID NOs: 10, 12, 6 and/or 8. In alternative embodiments, the homologous polypeptides have an amino acid sequence that differs by forty, thirty-five, thirty, twenty-five, twenty, or fifteen amino acids from amino acids 1 to 186 of SEQ ID NO: 2, or from the corresponding parts of SEQ ID NOs: 10, 12, 6 and/or 8.

In a particular embodiment, the polypeptides of the present invention comprise the amino acid sequence of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2, or the corresponding parts of any one of SEQ ID NOs: 12, 10, 6, and/or 8; or allelic variants thereof; or fragments thereof that have protease activity.

In another preferred embodiment, the polypeptides of the present invention consist of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2, or the corresponding parts of any one of SEQ ID NOs: 12, 10, 6, and/or 8; or allelic variants thereof or fragments thereof that have protease activity.

A fragment of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2, or the corresponding parts of any one of SEQ ID NOs: 10, 12, 8 and/or 6, is a polypeptide having one or more amino acids deleted from the amino and/or carboxyl terminus of these amino acid sequences. In one embodiment a fragment contains at least 75 amino acid residues, or at least 100 amino acid residues, or at least 125 amino acid residues, or at least 150 amino acid residues, or at least 160 amino acid residues, or at least 165 amino acid residues, or at least 170 amino acid residues, or at least 175 amino acid residues.

An allelic variant denotes any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene.

The present invention also relates to isolated polypeptides having protease activity and which are encoded by nucleic acid sequences which hybridize under very low, or low, or medium, or medium-high, or high, or very high stringency conditions with a nucleic acid probe which hybridizes under the same conditions with (a) nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1, (b) a subsequence of (a), or (c) a complementary strand of (a), or (b) (J. Sambrook, E.F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 2nd edition, Cold Spring Harbor, New York). In one particular embodiment the nucleic acid probe is selected from amongst the nucleic acid sequences of (a), (b), or (c) above. Nucleotides 726-1283 corresponding to the mature peptide encoding part of SEQ ID NO: 1 is a preferred probe. Other preferred probes are nucleotides 652-1206, 139-1206, or 82-1206 of SEQ ID NO: 9; nucleotides 638-1195, 134-1195, or 77-1195 of SEQ ID NO: 7; nucleotides 636-1193, 114-1193, or 57-1193 of SEQ ID NO: 5; and/or nucleotides 355-549, or 1-549 of SEQ ID NO: 11.

The subsequence of nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1, or any one of the other sequences specified in the last sentence hereinabove, may be at least 100 nucleotides, or in another embodiment at least 200 nucleotides. Moreover, the subsequence may encode a polypeptide fragment that has protease activity.

The nucleic acid sequences of nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1 or a subsequence thereof, as well as the amino acid sequences of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2 or a fragment thereof, as well as the corresponding parts of SEQ ID NOs: 5, 11, 9, 7 and SEQ ID NOs: 6, 12, 10, and 8, may be used to design a nucleic acid probe to identify and clone DNA encoding polypeptides having protease activity from strains of different genera or species according to methods well known in the art. In particular, such probes can be used for hybridization with the genomic DNA or cDNA of the genus or species of interest, following standard Southern blotting procedures, in order to identify and isolate the corresponding gene therein. Such probes can be considerably shorter than the entire sequence, but should be at least 15, preferably at least 25, and more preferably at least 35 nucleotides in length. Longer probes can also be used. Both DNA and RNA probes can be used. The probes are typically labeled for detecting the corresponding gene (for example, with ^{32}P, ^{3}H, ^{35}S, biotin, or avidin). Such probes are encompassed by the present invention.

Thus, a genomic DNA or cDNA library prepared from such other organisms may be screened for DNA that hybridizes with the probes described above and which encodes a polypeptide having protease activity. Genomic or other DNA from such other organisms may be separated by agarose or polyacrylamide gel electrophoresis, or other separation techniques. DNA from the libraries or the separated DNA may be transferred to and immobilized on nitrocellulose or other suitable carrier material. In order to identify a clone or DNA which is homologous with SEQ ID NO: 1 or a subsequence thereof, the carrier material is used in a Southern blot. For purposes of the present invention, hybridization indicates that the nucleic acid sequence hybridizes to a labelled nucleic acid probe corresponding to the nucleic acid sequence shown in SEQ ID NO: 1, its complementary strand, or a subsequence thereof, under very low to very high stringency conditions. Molecules to which the nucleic acid probe hybridizes under these conditions may be detected using X-ray film.

In a particular embodiment, the nucleic acid probe is a nucleic acid sequence which encodes amino acids -189 to 186, -170 to 186, or 1 to 186 of SEQ ID NO: 2, or subsequences thereof. In another embodiment, the nucleic acid probe is nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1, preferably nucleotides 726-1283 (the mature polypeptide coding region of SEQ ID NO: 1), or the corresponding parts of the other sequences referred to above. In another preferred embodiment, the nucleic acid probe is the nucleic acid sequence, or preferably the mature polypeptide coding region thereof, which is contained in the plasmid which is contained in *Escherichia coli* DSM 15509, DSM 16128 DSM 16052, or DSM 16051,wherein the nucleic acid sequence encodes a polypeptide having protease activity.

For long probes of at least 100 nucleotides in length, very low to very high stringency conditions are defined as prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 µg/ml sheared and denatured salmon sperm DNA, and either 25% formamide for very low and low stringencies, 35% formamide for medium and medium-high stringencies, or 50% formamide for high and very high stringencies, following standard Southern blotting procedures.

For long probes of at least 100 nucleotides in length, the carrier material is finally washed three times each for 15 minutes using 2 x SSC, 0.2% SDS preferably at least at 45°C (very low stringency), more preferably at least at 50°C (low stringency), more preferably at least at 55°C (medium stringency), more preferably at least at 60°C (medium-high stringency), even more preferably at least at 65°C (high stringency), and most preferably at least at 70°C (very high stringency).

For short probes about 15 nucleotides to about 70 nucleotides in length, stringency conditions are defined as prehybridization, hybridization, and washing post-hybridization at 5°C to 10°C below the calculated T_m using the calculation according to Bolton and McCarthy (1962, Proceedings of the National Academy of Sciences USA 48:1390) in 0.9 M NaCl, 0.09 M Tris-HCl pH 7.6, 6 mM EDTA, 0.5% NP-40, 1X Denhardt's solution, 1 mM sodium pyrophosphate, 1 mM sodium monobasic phosphate, 0.1 mM ATP, and 0.2 mg of yeast RNA per ml following standard Southern blotting procedures.

For short probes about 15 nucleotides to about 70 nucleotides in length, the carrier material is washed once in 6X SSC plus 0.1% SDS for 15 minutes and twice each for 15 minutes using 6X SSC at 5°C to 10°C below the calculated T_m.

The present invention also relates to variants of the polypeptide having an amino acid sequence of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2; and/or amino acids 1 to 185, 171 to 185, or 190 to 185 of SEQ ID NO: 10; amino acids 1 to 186, -168 to 186, or 187 to 186 of SEQ ID NO: 8; amino acids 1 to 186, 174 to 186, or 193 to 186 of SEQ ID NO: 6; amino acids 1 to 65, or 118 to 65 of SEQ ID NO: 12; comprising a substitution, deletion, and/or insertion of one or more amino acids.

The amino acid sequences of the variant polypeptides may differ from the amino acid indicated above by an insertion or deletion of one or more amino acid residues and/or the substitution of one or more amino acid residues by different amino acid residues. Preferably, amino acid changes are of a minor nature, that is conservative amino acid substitutions that do not significantly affect the folding and/or activity of the protein; small deletions, typically of one to about 30 amino acids; small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue; a small linker peptide of up to about 20-25 residues; or a small extension that facilitates purification by changing net charge or another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

Examples of conservative substitutions are within the group of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagine), hydrophobic amino acids (leucine, isoleucine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, threonine and methionine). Amino acid substitutions which do not generally alter the specific activity are known in the art and are described, for example, by H. Neurath and R.L. Hill, 1979, *In*, *The Proteins*, Academic Press, New York. The most commonly occurring exchanges are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, and Asp/Gly as well as these in reverse.

In a particular embodiment, the polypeptides of the invention and for use according to the invention are acid-stable. For the present purposes, the term acid-stable means that the residual activity after 2 hours of incubation at pH 3.0 and 37°C, is at least 20%, as compared to the residual activity of a corresponding sample incubated for 2 hours at pH 9.0 and 5°C. In a particular embodiment, the residual activity is at least 22%, 24%, 25% or at least 26%. In the alternative, the acid-stability definition refers to a residual activity of at least 50%, or 60%, or 70% when measured at pH 3.5 and 37°C, as compared to the residual activity of a corresponding sample incubated for 2 hours at pH 9.0 and 5°C. A suitable assay for determining acid-stability is the pH-stability assay of Example 1.

In another particular embodiment, the polypeptides of the invention and for use according to the invention have a relative activity at pH 7.0 of at least 0.2, 0.3, 0.4, or at least 0.5. The pH-profile test of Example 1 is used for the determination.

In still further particular embodiments, the polypeptides of the invention and for use according to the invention have i) a relative activity at 50°C of at least 0.4, 0.5, or at least 0.6; and/or ii) a relative activity at 60°C of at least 0.7, 0.8, or at least 0.9. The temperature-profile test of Example 1 is used for these determinations.

The polypeptide of the invention and for use according to the invention may be a bacterial or fungal polypeptide. The fungal polypeptide may be derived from a filamentous fungus or from a yeast.

In particular embodiments, the polypeptide of the invention is i) a fungal protease; ii) a protease derived from the phylum Ascomycota; iii) the subphylum Pezizomycotina; iv) the class Sordariomycetes; v) the order Sordariales; vi) the family Annulatascaceae; vii) the genus *Ascotaiwania* and/or *Brachysporiella* (*Brachysporiella* being the anamorphic (asexual) state of this fungus, and *Ascotaiwania* being the teleomorphic or sexual state); and/or viii) a protease derived from a strain of *Ascotaiwania* and/or *Brachysporiella*, for example *Ascotaiwania mitriformis*, *Ascotaiwania sawada*, *Brachysporiella gayana*, and *Brachysporiella sp.*, for example *Brachysporiella gayana* CGMCC 0865, such as a

polypeptide with the amino acid sequence of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2.

Other examples of polypeptides of the invention are derived from *Gliocladium sp.* CBS 114001, *Periconia sp.* CBS 114002, *Periconia sp.* CBS 114000, and *Curvularia lunata* CBS 114003.

In further particular embodiments, the polypeptides of the invention belong to either of the three sets of phyla, subphyla, classes, orders, families, or genera listed below:

Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; mitosporic Hypocreales, species of *Gliocladium*, such as *Gliocladium cibotii*, *Gliocladium deliquescens*, *Gliocladium flavofuscum*, and *Gliocladium viride*.

Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Halosphaeriales; Halosphaeriaceae; mitosporic Halosphaeriaceae, species of *Periconia*, such as *Periconia macrospinosa*, and *Periconia prolifica*.

Ascomycota; Pezizomycotina; Dothideomycetes; Pleosporales; Pleosporaceae; Cochliobolus; mitosporic Cochliobolus, species of *Curvularia*, such as *Curvularia affinis*, *Curvularia brachyspora*, *Curvularia clavata*, *Curvularia cymbopogonis*, *Curvularia eragrostidis*, *Curvularia gladioli*, *Curvularia gudauskasii*, *Curvularia heteropogonicola*, *Curvularia inaequalis*, *Curvularia intermedia*, *Curvularia oryzae*, and *Curvularia trifolii*.

In further particular embodiments, the polypeptide for use according to the invention is i) a polypeptide derived from a filamentous fungus; ii) of the phylum Ascomycota; iii) of the subphylum Pezizomycotina; iv) of the class Sordariomycetes; v) of the order Hypocreales or Sordariales; vi) of the families Annulatascaceae, Clavicipitaceae or Nectriaceae; e.g. mitosporic Clavicipitaceae (the anamorphic form), and/or vii) of the genus *Ascotaiwania*, *Brachysporiella*, or *Metarhizium*, such as the species *Metarhizium album*, *Metarhizium anisopliae*, *Metarhizium anisopliae* var. *acridum*, *Metarhizium anisopliae* var. *anisopliae*, *Metarhizium anisopliae* var. *frigidum*, *Metarhizium anisopliae* var. *lepidiotum*, *Metarhizium anisopliae* var. *major*, *Metarhizium cylindrosporae*, *Metarhizium flavoviride*, *Metarhizium flavoviride* var. *flavoviride*, *Metarhizium flavoviride* var. *minus*, *Metarhizium flavoviride* var. *novazealandicum*, *Metarhizium flavoviride* var. *pemphigum*, *Metarhizium flavoviride* var. *type E*, *Metarhizium sp. KACC 40230*, such as the polypeptide comprising amino acids 1 to 188, -168 to 188, or -186 to 188 of SEQ ID NO: 4; or the species *Ascotaiwania mitriformis*, *Ascotaiwania sawada*, *Brachysporiella gayana*, and *Brachysporiella sp.*, for example *Brachysporiella gayana* CGMCC 0865, such as a polypeptide with the amino acid sequence of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2.

The above taxonomy is mainly according to Ranghoo, V.M., Goh, T.K. & Hyde, K.D. 1999: New observations on *Monotosporella rhizoidea,* Mycosciences 40: 377-382; and Sivichai, S., Hywel-Jones, N. & J. E.B.G. 1998, Liginicolous freshwater Ascomycota from

Thailand: I. *Ascotaiwania sawada* and its anamorph state *Monotosporella,* Mycosciense 39: 307-311.

It will be understood that for the aforementioned species, the invention encompasses both the perfect and imperfect states, and other taxonomic equivalents, *e.g.*, anamorphs, regardless of the species name by which they are known. Those skilled in the art will readily recognize the identity of appropriate equivalents.

Strains of these species are readily accessible to the public in a number of culture collections, such as the American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

Furthermore, such polypeptides may be identified and obtained from other sources including microorganisms isolated from nature (*e.g.*, soil, composts, water, etc.) using the above-mentioned probes. Techniques for isolating microorganisms from natural habitats are well known in the art. The nucleic acid sequence may then be derived by similarly screening a genomic DNA or cDNA library of another microorganism. Once a nucleic acid sequence encoding a polypeptide has been detected with the probe(s), the sequence may be isolated or cloned by utilizing techniques which are known to those of ordinary skill in the art (see, *e.g.*, Sambrook *et al.*, 1989, *supra*).

As defined herein, an "isolated" polypeptide is a polypeptide which is essentially free of other polypeptides, *e.g.*, at least about 20% pure, preferably at least about 40% pure, more preferably about 60% pure, even more preferably about 80% pure, most preferably about 90% pure, and even most preferably about 95% pure, as determined by SDS-PAGE.

Polypeptides encoded by nucleic acid sequences of the present invention also include fused polypeptides or cleavable fusion polypeptides in which another polypeptide is fused at the N-terminus or the C-terminus of the polypeptide or fragment thereof. A fused polypeptide is produced by fusing a nucleic acid sequence (or a portion thereof) encoding another polypeptide to a nucleic acid sequence (or a portion thereof) of the present invention. Techniques for producing fusion polypeptides are known in the art, and include ligating the coding sequences encoding the polypeptides so that they are in frame and that expression of the fused polypeptide is under control of the same promoter(s) and terminator.

Nucleic Acid Sequences

The present invention also relates to isolated nucleic acid sequences that encode a polypeptide of the present invention. Particular nucleic acid sequences of the invention are nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1, 652-1206, 139-1206, or 82-1206 of SEQ ID NO: 9; nucleotides 638-1195, 134-1195, or 77-1195 of

SEQ ID NO: 7; nucleotides 636-1193, 114-1193, or 57-1193 of SEQ ID NO: 5; and/or nucleotides 355-549, or 1-549 of SEQ ID NO: 11, in particular the mature polypeptide encoding regions thereof. Other particular nucleic acid sequence of the invention are the sequences, preferably the mature polypeptide encoding regions thereof, which are contained in the plasmid that is contained in the deposited microorganism *Escherichia coli* strains DSM 15509, DSM 16128, DSM 16052, or DSM 16051. The present invention also encompasses nucleic acid sequences which encode a polypeptide having the amino acid sequence of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2; amino acids 1 to 185, -171 to 185, or -190 to 185 of SEQ ID NO: 10; amino acids 1 to 186, -168 to 186, or -187 to 186 of SEQ ID NO: 8; amino acids 1 to 186, -174 to 186, or -193 to 186 of SEQ ID NO: 6; and/or amino acids 1 to 65, or -118 to 65 of SEQ ID NO: 12; which differ from the corresponding parts of SEQ ID NOs: 1, 9, 7, 5, and 11 by virtue of the degeneracy of the genetic code. The present invention also relates to subsequences of SEQ ID NOs: 1, 5, 7, 9 and/or 11 which encode fragments of SEQ ID NOs: 2, 6, 8, 10 and/or 12 that have protease activity.

A subsequence is a nucleic acid sequence encompassed by the respective SEQ ID NO except that one or more nucleotides from the 5' and/or 3' end have been deleted. Preferably, a subsequence contains at least 150, 190 or at least 225 nucleotides, more preferably at least 300 nucleotides, even more preferably at least 375, 450, 500, 531, 600, 700, 800, 900, 1000, or 1100 nucleotides.

The present invention also relates to nucleotide sequences which have a degree of identity to nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1; nucleotides 652-1206, 139-1206, or 82-1206 of SEQ ID NO: 9; nucleotides 638-1195, 134-1195, or 77-1195 of SEQ ID NO: 7; nucleotides 636-1193, 114-1193, or 57-1193 of SEQ ID NO: 5; and/or nucleotides 355-549, or 1-549 of SEQ ID NO: 11; of at least 61%, 64%, 64%, 65%, and 57%, respectively. In particular embodiments, the degree of identity is at least 58%, 59%, 60%, 61%, 62%, 64%, 65%, 67%, 70%, 72%, 75, 77%, 80%, 82%, 85%, 87%, 90%, 92%, 95%, or at least 97%. In alternative embodiments, the degree of identity is at least 50%, 52%, 55%, 57%, or at least 60%. In a preferred embodiment, the degree of identity refers to the mature peptide parts of any one of SEQ ID NOs: 1, 5, 7, 9, and/or 11.

The present invention also relates to mutant nucleic acid sequences comprising at least one mutation in any one of the above specified nucleotide sequences, in which the mutant nucleic acid sequence encodes a polypeptide which (i) consists of the corresponding parts of the corresponding amino acid sequences, or (ii) is a variant of any of the sequences of (i), wherein the variant comprises a substitution, deletion, and/or insertion of one or more amino acids, or (iii) is an allelic variant of any of the sequences of (i), or (iv) is a fragment of any of the sequences of (i).

The techniques used to isolate or clone a nucleic acid sequence encoding a polypeptide are known in the art and include isolation from genomic DNA, preparation from cDNA, or a combination thereof. The cloning of the nucleic acid sequences of the present invention from such genomic DNA can be effected, *e.g.*, by using the well known polymerase chain reaction (PCR) or antibody screening of expression libraries to detect cloned DNA fragments with shared structural features. See, *e.g.*, Innis *et al.*, 1990, *PCR: A Guide to Methods and Application*, Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleic acid sequence-based amplification (NASBA) may be used. The nucleic acid sequence may be cloned from a strain of *Brachysporiella* (*Ascotaiwania*), or another or related organism and thus, for example, may be an allelic or species variant of the polypeptide encoding region of the nucleic acid sequence.

The term "isolated nucleic acid sequence" as used herein refers to a nucleic acid sequence which is essentially free of other nucleic acid sequences, *e.g.*, at least about 20% pure, preferably at least about 40% pure, more preferably at least about 60% pure, even more preferably at least about 80% pure, and most preferably at least about 90% pure as determined by agarose electrophoresis. For example, an isolated nucleic acid sequence can be obtained by standard cloning procedures used in genetic engineering to relocate the nucleic acid sequence from its natural location to a different site where it will be reproduced. The cloning procedures may involve excision and isolation of a desired nucleic acid fragment comprising the nucleic acid sequence encoding the polypeptide, insertion of the fragment into a vector molecule, and incorporation of the recombinant vector into a host cell where multiple copies or clones of the nucleic acid sequence will be replicated. The nucleic acid sequence may be of genomic, cDNA, RNA, semisynthetic, synthetic origin, or any combinations thereof.

Modification of a nucleic acid sequence encoding a polypeptide of the present invention may be necessary for the synthesis of polypeptides substantially similar to the polypeptide. The term "substantially similar" to the polypeptide refers to non-naturally occurring forms of the polypeptide. These polypeptides may differ in some engineered way from the polypeptide isolated from its native source, *e.g.*, variants that differ in specific activity, thermostability, pH optimum, allergenicity, or the like. The variant sequence may be constructed on the basis of the nucleic acid sequence presented as the polypeptide encoding part of SEQ ID NO: 1, 5, 7, 9 and/or 11, e.g., a subsequence thereof, and/or by introduction of nucleotide substitutions which do not give rise to another amino acid sequence of the polypeptide encoded by the nucleic acid sequence, but which correspond to the codon usage of the host organism intended for production of the protease, or by introduction of nucleotide substitutions which may give rise to a different amino acid

sequence. For a general description of nucleotide substitution, see, *e.g.*, Ford *et al.*, 1991, *Protein Expression and Purification* 2: 95-107. Low-allergenic polypeptides can e.g. be prepared as described above.

It will be apparent to those skilled in the art that such substitutions can be made outside the regions critical to the function of the molecule and still result in an active polypeptide. Amino acid residues essential to the activity of the polypeptide encoded by the isolated nucleic acid sequence of the invention, and therefore preferably not subject to substitution, may be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (see, *e.g.*, Cunningham and Wells, 1989, *Science* 244: 1081-1085). In the latter technique, mutations are introduced at every positively charged residue in the molecule, and the resultant mutant molecules are tested for protease activity to identify amino acid residues that are critical to the activity of the molecule. Sites of substrate-protease interaction can also be determined by analysis of the three-dimensional structure as determined by such techniques as nuclear magnetic resonance analysis, crystallography or photoaffinity labelling (see, *e.g.*, de Vos *et al.*, 1992, *Science* 255: 306-312; Smith *et al.*, 1992, *Journal of Molecular Biology* 224: 899-904; Wlodaver *et al.*, 1992, *FEBS Letters* 309: 59-64).

The present invention also relates to isolated nucleic acid sequences encoding a polypeptide of the present invention, which hybridize under very low stringency conditions, preferably low stringency conditions, more preferably medium stringency conditions, more preferably medium-high stringency conditions, even more preferably high stringency conditions, and most preferably very high stringency conditions with a nucleic acid probe which hybridizes under the same conditions with the nucleic acid sequence of any one of SEQ ID NOs: 1, 5, 7, 9, or 11, or their complementary strands; or allelic variants and subsequences thereof (Sambrook *et al.*, 1989, *supra*), as defined herein.

The present invention also relates to isolated nucleic acid sequences produced by (a) hybridizing a DNA under very low, low, medium, medium-high, high, or very high stringency conditions with (i) nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1; nucleotides 652-1206, 139-1206, or 82-1206 of SEQ ID NO: 9; nucleotides 638-1195, 134-1195, or 77-1195 of SEQ ID NO: 7; nucleotides 636-1193, 114-1193, or 57-1193 of SEQ ID NO: 5; and/or nucleotides 355-549, or 1-549 of SEQ ID NO: 11; (ii) a subsequence of (i), or (iii) a complementary strand of (i), or (ii); and (b) isolating the nucleic acid sequence. The subsequence is preferably a sequence of at least 100 nucleotides such as a sequence that encodes a polypeptide fragment which has protease activity.

Methods for Producing Mutant Nucleic Acid Sequences

The present invention further relates to methods for producing a mutant nucleic acid sequence, comprising introducing at least one mutation into the mature polypeptide coding sequence of SEQ ID NO: 1, 5, 7, 9 or 11, or a subsequence thereof, wherein the mutant nucleic acid sequence encodes a polypeptide which consists of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2 or the corresponding parts of SEQ ID NOs. 6, 8, 10 or 12; or a fragment thereof which has protease activity.

The introduction of a mutation into the nucleic acid sequence to exchange one nucleotide for another nucleotide may be accomplished by site-directed mutagenesis using any of the methods known in the art. Particularly useful is the procedure that utilizes a supercoiled, double stranded DNA vector with an insert of interest and two synthetic primers containing the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector, extend during temperature cycling by means of *Pfu* DNA polymerase. On incorporation of the primers, a mutated plasmid containing staggered nicks is generated. Following temperature cycling, the product is treated with *Dpn*I which is specific for methylated and hemimethylated DNA to digest the parental DNA template and to select for mutation-containing synthesized DNA. Other procedures known in the art may also be used.

Nucleic Acid Constructs

The present invention also relates to nucleic acid constructs comprising a nucleic acid sequence of the present invention operably linked to one or more control sequences that direct the expression of the coding sequence in a suitable host cell under conditions compatible with the control sequences. Expression will be understood to include any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

"Nucleic acid construct" is defined herein as a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or which has been modified to contain segments of nucleic acid combined and juxtaposed in a manner that would not otherwise exist in nature. The term nucleic acid construct is synonymous with the term expression cassette when the nucleic acid construct contains all the control sequences required for expression of a coding sequence of the present invention. The term "coding sequence" is defined herein as a nucleic acid sequence that directly specifies the amino acid sequence of its protein product. The boundaries of the coding sequence are generally determined by a ribosome binding site (prokaryotes) or by the ATG start codon (eukaryotes) located just upstream of the open reading frame at the 5' end of the mRNA and a transcription terminator sequence located just downstream of the open reading frame at the

3' end of the mRNA. A coding sequence can include, but is not limited to, DNA, cDNA, and recombinant nucleic acid sequences.

An isolated nucleic acid sequence encoding a polypeptide of the present invention may be manipulated in a variety of ways to provide for expression of the polypeptide. Manipulation of the nucleic acid sequence prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

The term "control sequences" is defined herein to include all components that are necessary or advantageous for the expression of a polypeptide of the present invention. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide. The term "operably linked" is defined herein as a configuration in which a control sequence is appropriately placed at a position relative to the coding sequence of the DNA sequence such that the control sequence directs the expression of a polypeptide.

The control sequence may be an appropriate promoter sequence, a nucleic acid sequence that is recognized by a host cell for expression of the nucleic acid sequence. The promoter sequence contains transcriptional control sequences that mediate the expression of the polypeptide. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention in a filamentous fungal host cell are promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (*glaA*), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (WO 96/00787), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof.

In a yeast host, useful promoters are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are described by Romanos *et al.*, 1992, *Yeast* 8: 423-488.

The control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice may be used in the present invention.

Preferred terminators for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease.

Preferred terminators for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are described by Romanos *et al.*, 1992, *supra*.

The control sequence may also be a suitable leader sequence, a nontranslated region of an mRNA which is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used in the present invention.

Preferred leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase.

Suitable leaders for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention.

Preferred polyadenylation sequences for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase,

Aspergillus nidulans anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase.

Useful polyadenylation sequences for yeast host cells are described by Guo and Sherman, 1995, *Molecular Cellular Biology* 15: 5983-5990.

The control sequence may also be a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region which encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region which is foreign to the coding sequence. The foreign signal peptide coding region may be required where the coding sequence does not naturally contain a signal peptide coding region. Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to enhance secretion of the polypeptide. However, any signal peptide coding region which directs the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.

Effective signal peptide coding regions for filamentous fungal host cells are the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase.

In a preferred embodiment, the signal peptide coding region is nucleotides 159-215 of SEQ ID NO: 1 which encode amino acids -189 to -171 of SEQ ID NO: 2, or nucleotides 1-54 of SEQ ID NO: 3 which encode amino acids -186 to -169 of SEQ ID NO: 4. The signal peptide coding regions of SEQ ID NOs: 5, 7, 9 and 11 are other preferred signal peptide coding regions.

Useful signal peptides for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are described by Romanos *et al.*, 1992, *supra*.

The control sequence may also be a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained, e.g., from the genes for *Bacillus subtilis* alkaline protease (*aprE*), *Bacillus subtilis* neutral protease (*nprT*), *Saccharomyces cerevisiae* alpha-factor,

Rhizomucor miehei aspartic proteinase, and *Myceliophthora thermophila* laccase (WO 95/33836).

In a preferred embodiment, the propeptide coding region is nucleotides 216-725 of SEQ ID NO: 1 which encode amino acids -170 to -1 of SEQ ID NO: 2, or nucleotides 55-558 of SEQ ID NO: 3 which encode amino acids -168 to -1 of SEQ ID NO: 4. Other preferred propeptide coding regions are the corresponding regions of SEQ ID NOs: 5, 7, 9 and 11.

Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

It may also be desirable to add regulatory sequences which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Regulatory systems in prokaryotic systems include the *lac*, *tac*, and *trp* operator systems. In yeast, the ADH2 system or GAL1 system may be used. In filamentous fungi, the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter may be used as regulatory sequences. Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene which is amplified in the presence of methotrexate, and the metallothionein genes which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the polypeptide would be operably linked with the regulatory sequence.

Expression Vectors

The present invention also relates to recombinant expression vectors comprising a nucleic acid sequence of the present invention, a promoter, and transcriptional and translational stop signals. The various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, the nucleic acid sequence of the present invention may be expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

The recombinant expression vector may be any vector (*e.g.*, a plasmid or virus) which can be conveniently subjected to recombinant DNA procedures and can bring about

the expression of the nucleic acid sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

The vector may be an autonomously replicating vector, *i.e.*, a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used.

The vectors of the present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*. Suitable markers for yeast host cells are ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host cell include, but are not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferase), *bar* (phosphinothricin acetyltransferase), *hygB* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenyltransferase), *trpC* (anthranilate synthase), as well as equivalents thereof. Preferred for use in an *Aspergillus* cell are the *amdS* and *pyrG* genes of *Aspergillus nidulans* or *Aspergillus oryzae* and the *bar* gene of *Streptomyces hygroscopicus*.

The vectors of the present invention preferably contain an element(s) that permits stable integration of the vector into the host cell's genome or autonomous replication of the vector in the cell independently of the genome.

For integration into the host cell genome, the vector may rely on the nucleic acid sequence encoding the polypeptide or any other element of the vector for stable integration of the vector into the genome by homologous or nonhomologous recombination. Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 1,500 base pairs, preferably 400 to 1,500 base pairs, and most preferably 800 to 1,500 base pairs, which are highly homologous with the corresponding

target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are the origins of replication of plasmids pBR322, pUC19, pACYC177, and pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, and pAMß1 permitting replication in *Bacillus*. Examples of origins of replication for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6. The origin of replication may be one having a mutation which makes its functioning temperature-sensitive in the host cell (see, *e.g.*, Ehrlich, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1433).

More than one copy of a nucleic acid sequence of the present invention may be inserted into the host cell to increase production of the gene product. An increase in the copy number of the nucleic acid sequence can be obtained by integrating at least one additional copy of the sequence into the host cell genome or by including an amplifiable selectable marker gene with the nucleic acid sequence where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the nucleic acid sequence, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

The procedures used to ligate the elements described above to construct the recombinant expression vectors of the present invention are well known to one skilled in the art (see, *e.g.*, Sambrook *et al.*, 1989, *supra*).

The protease may also be co-expressed together with at least one other enzyme of interest for animal feed, such as phytase (EC 3.1.3.8 or 3.1.3.26); xylanase (EC 3.2.1.8); galactanase (EC 3.2.1.89); alpha-galactosidase (EC 3.2.1.22); protease (EC 3.4.-.-), phospholipase A1 (EC 3.1.1.32); phospholipase A2 (EC 3.1.1.4); lysophospholipase (EC 3.1.1.5); phospholipase C (3.1.4.3); phospholipase D (EC 3.1.4.4); and/or beta-glucanase (EC 3.2.1.4 or EC 3.2.1.6).

The enzymes may be co-expressed from different vectors, from one vector, or using a mixture of both techniques. When using different vectors, the vectors may have different selectable markers, and different origins of replication. When using only one vector, the genes can be expressed from one or more promoters. If cloned under the regulation of one promoter (di- or multi-cistronic), the order in which the genes are cloned may affect the expression levels of the proteins. The protease may also be expressed as a fusion protein, i.e. that the gene encoding the protease has been fused in frame to the gene encoding

another protein. This protein may be another enzyme or a functional domain from another enzyme.

Host Cells

The present invention also relates to recombinant host cells, comprising a nucleic acid sequence of the invention, which are advantageously used in the recombinant production of the polypeptides. A vector comprising a nucleic acid sequence of the present invention is introduced into a host cell so that the vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector as described earlier. The term "host cell" encompasses any progeny of a parent cell that is not identical to the parent cell due to mutations that occur during replication. The choice of a host cell will to a large extent depend upon the gene encoding the polypeptide and its source.

The host cell may be a unicellular microorganism, *e.g.*, a prokaryote, or a non-unicellular microorganism, *e.g.*, a eukaryote.

Useful unicellular cells are bacterial cells such as gram positive bacteria including, but not limited to, a *Bacillus* cell, or a *Streptomyces* cell, or cells of lactic acid bacteria; or gram negative bacteria such as *E. coli* and *Pseudomonas* sp. Lactic acid bacteria include, but are not limited to, species of the genera *Lactococcus*, *Lactobacillus*, *Leuconostoc*, *Streptococcus*, *Pediococcus*, and *Enterococcus*.

The introduction of a vector into a bacterial host cell may, for instance, be effected by protoplast transformation (see, *e.g.*, Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115), using competent cells (see, *e.g.*, Young and Spizizin, 1961, *Journal of Bacteriology* 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221), electroporation (see, *e.g.*, Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), or conjugation (see, *e.g.*, Koehler and Thorne, 1987, *Journal of Bacteriology* 169: 5771-5278).

The host cell may be a eukaryote, such as a non-human animal cell, an insect cell, a plant cell, or a fungal cell.

In one particular embodiment, the host cell is a fungal cell. "Fungi" as used herein includes the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota (as defined by Hawksworth et al., In Ainsworth and Bisby's Dictionary of The Fungi, 8th edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited in Hawksworth et al., 1995, supra, page 171) and all mitosporic fungi (Hawksworth et al., 1995, supra).

In another particular embodiment, the fungal host cell is a yeast cell. "Yeast" as used herein includes ascosporogenous yeast (Endomycetales), basidiosporogenous yeast, and yeast belonging to the Fungi Imperfecti (Blastomycetes). Since the classification of yeast

may change in the future, for the purposes of this invention, yeast shall be defined as described in Biology and Activities of Yeast (Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980).

The yeast host cell may be a *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* cell.

The fungal host cell may be a filamentous fungal cell. "Filamentous fungi" include all filamentous forms of the subdivision Eumycota and Oomycota (as defined by Hawksworth *et al.*, 1995, *supra*). The filamentous fungi are characterized by a mycelial wall composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In contrast, vegetative growth by yeasts such as *Saccharomyces cerevisiae* is by budding of a unicellular thallus and carbon catabolism may be fermentative.

Examples of filamentous fungal host cells are cells of species of, but not limited to, *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium*, or *Trichoderma*.

Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* host cells are described in EP 238 023 and Yelton et al., 1984, Proceedings of the National Academy of Sciences USA 81: 1470-1474. Suitable methods for transforming Fusarium species are described by Malardier et al., 1989, Gene 78: 147-156 and WO 96/00787. Yeast may be transformed using the procedures described by Becker and Guarente, In Abelson, J.N. and Simon, M.I., editors, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito et al., 1983, Journal of Bacteriology 153: 163; and Hinnen et al., 1978, Proceedings of the National Academy of Sciences USA 75: 1920.

Methods of Production

The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a strain, which in its wild-type form is capable of producing the polypeptide; and (b) recovering the polypeptide. Preferably, the strain is of the genus *Brachysporiella*, such as *Brachysporiella sp.* CGMCC 0865, of the genus *Gliocladium*, such as *Gliocladium sp.* CBS 114001, of the genus *Periconia*, such as *Periconia sp.* CBS 114002 or CBS 114000, or of the genus, such as *Curvularia sp.* CBS 114003.

The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a host cell under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a host cell under conditions conducive for production of the polypeptide, wherein the host cell comprises a mutant nucleic acid sequence comprising at least one mutation in nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1, or in the corresponding parts of SEQ ID NOs: 5, 7, 9 or 11, in which the mutant nucleic acid sequence encodes a polypeptide which (i) consists of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2, or the corresponding parts of SEQ ID NOs: 6, 8, 10 or 12; or (ii) is a variant of any of the sequences of (i), wherein the variant comprises a substitution, deletion, and/or insertion of one or more amino acids, or (iii) is an allelic variant of any of the sequences of (i), or (iv) is a fragment of any of the sequences of (i).

In the production methods of the present invention, the cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the polypeptide to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (*e.g.*, in catalogues of the American Type Culture Collection). If the polypeptide is secreted into the nutrient medium, the polypeptide can be recovered directly from the medium. If the polypeptide is not secreted, it can be recovered from cell lysates.

The polypeptides may be detected using methods known in the art that are specific for the polypeptides. These detection methods may include use of specific antibodies, formation of a product, or disappearance of a substrate. For example, a protease assay may be used to determine the activity of the polypeptide as described herein.

The resulting polypeptide may be recovered by methods known in the art. For example, the polypeptide may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation.

The polypeptides of the present invention may be purified by a variety of procedures known in the art including, but not limited to, chromatography (*e.g.*, ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing), differential solubility (e.g., ammonium sulfate precipitation), SDS-PAGE, or extraction (see, e.g., Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

Plants

The present invention also relates to a transgenic plant, plant part, or plant cell which has been transformed with a nucleic acid sequence encoding a polypeptide having protease activity of the present invention so as to express and produce the polypeptide in recoverable quantities. The polypeptide may be recovered from the plant or plant part. Alternatively, the plant or plant part containing the recombinant polypeptide may be used as such for improving the quality of a food or feed, *e.g.*, improving nutritional value, palatability, and rheological properties, or to destroy an antinutritive factor.

In a particular embodiment, the polypeptide is targeted to the endosperm storage vacuoles in seeds. This can be obtained by synthesizing it as a precursor with a suitable signal peptide, see Horvath et al in PNAS, Feb. 15, 2000, vol. 97, no. 4, p. 1914-1919.

The transgenic plant can be dicotyledonous (a dicot) or monocotyledonous (a monocot) or engineered variants thereof. Examples of monocot plants are grasses, such as meadow grass (blue grass, Poa), forage grass such as Festuca, Lolium, temperate grass, such as Agrostis, and cereals, *e.g.*, wheat, oats, rye, barley, rice, sorghum, and maize (corn). Examples of dicot plants are tobacco, legumes, such as lupins, potato, sugar beet, pea, bean and soybean, and cruciferous plants (family Brassicaceae), such as cauliflower, rape seed, and the closely related model organism *Arabidopsis thaliana*. Low-phytate plants as described e.g. in US patent no. 5,689,054 and US patent no. 6,111,168 are examples of engineered plants.

Examples of plant parts are stem, callus, leaves, root, fruits, seeds, and tubers, as well as the individual tissues comprising these parts, e.g. epidermis, mesophyll, parenchyma, vascular tissues, meristems. Also specific plant cell compartments, such as chloroplast, apoplast, mitochondria, vacuole, peroxisomes, and cytoplasm are considered to be a plant part. Furthermore, any plant cell, whatever the tissue origin, is considered to be a plant part. Likewise, plant parts such as specific tissues and cells isolated to facilitate the utilisation of the invention are also considered plant parts, e.g. embryos, endosperms, aleurone and seed coats.

Also included within the scope of the present invention are the progeny of such plants, plant parts and plant cells.

The transgenic plant or plant cell expressing a polypeptide of the present invention may be constructed in accordance with methods known in the art. Briefly, the plant or plant cell is constructed by incorporating one or more expression constructs encoding a polypeptide of the present invention into the plant host genome and propagating the resulting modified plant or plant cell into a transgenic plant or plant cell.

Conveniently, the expression construct is a nucleic acid construct which comprises a nucleic acid sequence encoding a polypeptide of the present invention operably linked with appropriate regulatory sequences required for expression of the nucleic acid sequence in the plant or plant part of choice. Furthermore, the expression construct may comprise a selectable marker useful for identifying host cells into which the expression construct has been integrated and DNA sequences necessary for introduction of the construct into the plant in question (the latter depends on the DNA introduction method to be used).

The choice of regulatory sequences, such as promoter and terminator sequences and optionally signal or transit sequences are determined, for example, on the basis of when, where, and how the polypeptide is desired to be expressed. For instance, the expression of the gene encoding a polypeptide of the present invention may be constitutive or inducible, or may be developmental, stage or tissue specific, and the gene product may be targeted to a specific tissue or plant part such as seeds or leaves. Regulatory sequences are, for example, described by Tague *et al.*, 1988, *Plant Physiology* 86: 506.

For constitutive expression, the following may be used: The 35S-CaMV promoter may be used (Franck *et al.*, 1980, *Cell* 21: 285-294), the maize ubiquitin 1 (Christensen AH, Sharrock RA and Quail 1992. Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation), or the rice actin 1 promoter (Plant Mo. Biol. 18, 675-689.; Zhang W, McElroy D. and Wu R 1991, Analysis of rice Act1 5' region activity in transgenic rice plants. Plant Cell 3, 1155-1165). Organ-specific promoters may be, for example, a promoter from storage sink tissues such as seeds, potato tubers, and fruits (Edwards & Coruzzi, 1990, *Ann. Rev. Genet.* 24: 275-303), or from metabolic sink tissues such as meristems (Ito *et al.*, 1994, *Plant Mol. Biol.* 24: 863-878), a seed specific promoter such as the glutelin, prolamin, globulin, or albumin promoter from rice (Wu *et al.*, 1998, *Plant and Cell Physiology* 39: 885-889), a *Vicia faba* promoter from the legumin B4 and the unknown seed protein gene from *Vicia faba* (Conrad *et al.*, 1998, *Journal of Plant Physiology* 152: 708-711), a promoter from a seed oil body protein (Chen *et al.*, 1998, Plant and Cell Physiology 39: 935-941), the storage protein napA promoter from *Brassica napus*, or any other seed specific promoter known in the art, *e.g.*, as described in WO 91/14772. Furthermore, the promoter may be a leaf specific promoter such as the *rbcs* promoter from rice or tomato (Kyozuka *et al.*, 1993, *Plant Physiology* 102: 991-1000, the chlorella virus adenine methyltransferase gene promoter (Mitra and Higgins, 1994, *Plant Molecular Biology* 26: 85-93), or the *aldP* gene promoter from rice (Kagaya *et al.*, 1995, *Molecular and General Genetics* 248: 668-674), or a wound inducible promoter such as the potato pin2 promoter (Xu *et al.*, 1993, *Plant Molecular Biology* 22: 573-588). Likewise, the promoter may be inducible by abiotic treatments such as temperature, drought or alterations in salinity or inducible by

exogenously applied substances that activate the promoter, e.g. ethanol, oestrogens, plant hormones like ethylene, abscisic acid, gibberellic acid, and/or heavy metals.

A promoter enhancer element may also be used to achieve higher expression of the protease in the plant. For instance, the promoter enhancer element may be an intron which is placed between the promoter and the nucleotide sequence encoding a polypeptide of the present invention. For instance, Xu *et al.*, 1993, *supra* disclose the use of the first intron of the rice actin 1 gene to enhance expression.

Still further, the codon usage may be optimized for the plant species in question to improve expression (see Horvath et al referred to above).

The selectable marker gene and any other parts of the expression construct may be chosen from those available in the art.

The nucleic acid construct is incorporated into the plant genome according to conventional techniques known in the art, including *Agrobacterium*-mediated transformation, virus-mediated transformation, microinjection, particle bombardment, biolistic transformation, and electroporation (Gasser *et al.*, 1990, *Science* 244: 1293; Potrykus, 1990, *Bio/Technology* 8: 535; Shimamoto *et al.*, 1989, *Nature* 338: 274).

Presently, *Agrobacterium tumefaciens*-mediated gene transfer is the method of choice for generating transgenic dicots (for a review, see Hooykas and Schilperoort, 1992, *Plant Molecular Biology* 19: 15-38), and it can also be used for transforming monocots, although other transformation methods are generally preferred for these plants. Presently, the method of choice for generating transgenic monocots, supplementing the *Agrobacterium* approach, is particle bombardment (microscopic gold or tungsten particles coated with the transforming DNA) of embryonic calli or developing embryos (Christou, 1992, *Plant Journal* 2: 275-281; Shimamoto, 1994, *Current Opinion Biotechnology* 5: 158-162; Vasil *et al.*, 1992, *Bio/Technology* 10: 667-674). An alternative method for transformation of monocots is based on protoplast transformation as described by Omirulleh *et al.*, 1993, *Plant Molecular Biology* 21: 415-428.

Following transformation, the transformants having incorporated therein the expression construct are selected and regenerated into whole plants according to methods well-known in the art.

The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a transgenic plant or a plant cell comprising a nucleic acid sequence encoding a polypeptide having protease activity of the present invention under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

Animals

The present invention also relates to a transgenic, non-human animal and products or elements thereof, examples of which are body fluids such as milk and blood, organs, flesh, and animal cells. Techniques for expressing proteins, e.g. in mammalian cells, are known in the art, see e.g. the handbook Protein Expression: A Practical Approach, Higgins and Hames (eds), Oxford University Press (1999), and the three other handbooks in this series relating to Gene Transcription, RNA processing, and Post-translational Processing. Generally speaking, to prepare a transgenic animal, selected cells of a selected animal are transformed with a nucleic acid sequence encoding a polypeptide having protease activity of the present invention so as to express and produce the polypeptide. The polypeptide may be recovered from the animal, e.g. from the milk of female animals, or the polypeptide may be expressed to the benefit of the animal itself, e.g. to assist the animal's digestion. Examples of animals are mentioned below in the section headed Animal Feed.

To produce a transgenic animal with a view to recovering the protease from the milk of the animal, a gene encoding the protease may be inserted into the fertilized eggs of an animal in question, e.g. by use of a transgene expression vector which comprises a suitable milk protein promoter, and the gene encoding the protease. The transgene expression vector is microinjected into fertilized eggs, and preferably permanently integrated into the chromosome. Once the egg begins to grow and divide, the potential embryo is implanted into a surrogate mother, and animals carrying the transgene are identified. The resulting animal can then be multiplied by conventional breeding. The polypeptide may be purified from the animal's milk, see e.g. Meade, H.M. et al (1999): Expression of recombinant proteins in the milk of transgenic animals, Gene expression systems: Using nature for the art of expression. J. M. Fernandez and J. P. Hoeffler (eds.), Academic Press.

In the alternative, in order to produce a transgenic non-human animal that carries in the genome of its somatic and/or germ cells a nucleic acid sequence including a heterologous transgene construct including a transgene encoding the protease, the transgene may be operably linked to a first regulatory sequence for salivary gland specific expression of the protease, as disclosed in WO 2000064247.

Compositions

In a still further aspect, the present invention relates to compositions comprising a polypeptide of the present invention.

The polypeptide compositions may be prepared in accordance with methods known in the art and may be in the form of a liquid or a dry composition. For instance, the polypeptide composition may be in the form of a granulate or a microgranulate. The polypeptide to be included in the composition may be stabilized in accordance with methods known in the art.

Examples are given below of preferred uses of the polypeptides or polypeptide compositions of the invention.

Animal Feed

The present invention is also directed to methods for using the *Brachysporiella* protease and homologous proteases in animal feed, as well as to feed compositions and feed additives comprising the polypeptides of the invention. The mature peptide part of the *Brachysporiella* protease corresponds to amino acids 1-186 of SEQ ID NO: 2. An example of a homologous protease is the *Metarhizium* protease, the mature peptide part of which is amino acids 1-188 of SEQ ID NO: 4. Other examples are amino acids 1 to 185 of SEQ ID NO: 10, amino acids 1 to 186 of SEQ ID NO: 8, amino acids 1 to 186 of SEQ ID NO: 6, and a mature protease comprising amino acids 1 to 65 of SEQ ID NO: 12.

The term animal includes all animals, including human beings. Examples of animals are non-ruminants, and ruminants, such as sheep, goats, horses, and cattle, e.g. beef cattle, cows, and young calves. In a particular embodiment, the animal is a non-ruminant animal. Non-ruminant animals include mono-gastric animals, e.g. pigs or swine (including, but not limited to, piglets, growing pigs, and sows); poultry such as turkeys, ducks and chicken (including but not limited to broiler chicks, layers); young calves; and fish (including but not limited to salmon, trout, tilapia, catfish and carps); and crustaceans (including but not limited to shrimps and prawns).

The term feed or feed composition means any compound, preparation, mixture, or composition suitable for, or intended for intake by an animal.

In the use according to the invention the protease can be fed to the animal before, after, or simultaneously with the diet. The latter is preferred.

In a particular embodiment, the protease, in the form in which it is added to the feed, or when being included in a feed additive, is well-defined. Well-defined means that the protease preparation is at least 50% pure as determined by Size-exclusion chromatography (see Example 12 of WO 01/58275). In other particular embodiments the protease preparation is at least 60, 70, 80, 85, 88, 90, 92, 94, or at least 95% pure as determined by this method.

A well-defined protease preparation is advantageous. For instance, it is much easier to dose correctly to the feed a protease that is essentially free from interfering or contaminating other proteases. The term dose correctly refers in particular to the objective of obtaining consistent and constant results, and the capability of optimising dosage based upon the desired effect.

For the use in animal feed, however, the protease need not be that pure; it may e.g. include other enzymes, in which case it could be termed a protease preparation.

The protease preparation can be (a) added directly to the feed (or used directly in a treatment process of vegetable proteins), or (b) it can be used in the production of one or more intermediate compositions such as feed additives or premixes that is subsequently added to the feed (or used in a treatment process). The degree of purity described above refers to the purity of the original protease preparation, whether used according to (a) or (b) above.

Protease preparations with purities of this order of magnitude are in particular obtainable using recombinant methods of production, whereas they are not so easily obtained and also subject to a much higher batch-to-batch variation when the protease is produced by traditional fermentation methods.

Such protease preparation may of course be mixed with other enzymes.

In a particular embodiment, the protease for use according to the invention is capable of solubilising vegetable proteins. A suitable assay for determining solubilised protein is disclosed in Example 2.

The term vegetable proteins as used herein refers to any compound, composition, preparation or mixture that includes at least one protein derived from or originating from a vegetable, including modified proteins and protein-derivatives. In particular embodiments, the protein content of the vegetable proteins is at least 10, 20, 30, 40, 50, or 60% (w/w).

Vegetable proteins may be derived from vegetable protein sources, such as legumes and cereals, for example materials from plants of the families *Fabaceae* (*Leguminosae*), *Cruciferaceae*, *Chenopodiaceae*, and *Poaceae*, such as soy bean meal, lupin meal and rapeseed meal.

In a particular embodiment, the vegetable protein source is material from one or more plants of the family *Fabaceae*, e.g. soybean, lupine, pea, or bean.

In another particular embodiment, the vegetable protein source is material from one or more plants of the family *Chenopodiaceae*, e.g. beet, sugar beet, spinach or quinoa.

Other examples of vegetable protein sources are rapeseed, and cabbage.

Soybean is a preferred vegetable protein source.

Other examples of vegetable protein sources are cereals such as barley, wheat, rye, oat, maize (corn), rice, and sorghum.

The treatment according to the invention of vegetable proteins with at least one protease of the invention results in an increased solubilisation of vegetable proteins.

The following are examples of % solubilised protein obtainable using the proteases of the invention: At least 101%, 102%, 103%, or 104%, relative to a blank. The percentage of solubilised protein is determined using the *in vitro* model of Example 2.

The term solubilisation of proteins basically means bringing protein(s) into solution. Such solubilisation may be due to protease-mediated release of protein from other

components of the usually complex natural compositions such as feed. Solubilisation can be measured as an increase in the amount of soluble proteins, by reference to a blank sample with no protease treatment.

In a further particular embodiment, the protease for use according to the invention is capable of increasing the amount of digestible vegetable proteins. A suitable assay for determining digestible protein is disclosed in Example 2. The following are examples of % digested or digestible protein obtainable using the proteases of the invention: At least 101%, 102%, 103%, 104%, 105%, 106%, or 107%, relative to a blank. The percentage of digested or digestible protein is determined using the *in vitro* model of Example 2.

In a still further particular embodiment, the protease for use according to the invention is capable of increasing the Degree of Hydrolysis (DH) of vegetable proteins. A suitable assay for determining DH is disclosed in Example 2. The following are examples of Degree of Hydrolysis increase: At least 101%, 102%, 103%, 104%, 105%, 106%, or 107%, relative to a blank. The DH is determined using the *in vitro* model of Example 2.

In a particular embodiment of a (pre-) treatment process of the invention, the protease(s) in question is affecting (or acting on, or exerting its solubilising influence on) the vegetable proteins or protein sources. To achieve this, the vegetable protein or protein source is typically suspended in a solvent, eg an aqueous solvent such as water, and the pH and temperature values are adjusted paying due regard to the characteristics of the enzyme in question. For example, the treatment may take place at a pH-value at which the activity of the actual protease is at least Likewise, for example, the treatment may take place at a temperature at which the activity of the actual protease is at least 40%, 50%, 60%, 70%, 80% or at least 90%. The above percentage activity indications are relative to the maximum activities. The enzymatic reaction is continued until the desired result is achieved, following which it may or may not be stopped by inactivating the enzyme, e.g. by a heat-treatment step.

In another particular embodiment of a treatment process of the invention, the protease action is sustained, meaning e.g. that the protease is added to the vegetable proteins or protein sources, but its solubilising influence is so to speak not switched on until later when desired, once suitable solubilising conditions are established, or once any enzyme inhibitors are inactivated, or whatever other means could have been applied to postpone the action of the enzyme.

In one embodiment the treatment is a pre-treatment of animal feed or vegetable proteins for use in animal feed.

The term improving the nutritional value of an animal feed means improving the availability of the proteins, thereby leading to increased protein extraction, higher protein yields, and/or improved protein utilisation. The nutritional value of the feed is therefore

increased, and the growth rate and/or weight gain and/or feed conversion (i.e. the weight of ingested feed relative to weight gain) of the animal is/are improved.

The protease can be added to the feed in any form, be it as a relatively pure protease, or in admixture with other components intended for addition to animal feed, i.e. in the form of animal feed additives, such as the so-called pre-mixes for animal feed.

In a further aspect the present invention relates to compositions for use in animal feed, such as animal feed, and animal feed additives, e.g. premixes.

Apart from the protease of the invention, the animal feed additives of the invention contain at least one fat-soluble vitamin, and/or at least one water soluble vitamin, and/or at least one trace mineral. The feed additive may also contain at least one macro mineral

Further, optional, feed-additive ingredients are colouring agents, aroma compounds, stabilisers, antimicrobial peptides, including antifungal polypeptides, and/or at least one other enzyme selected from amongst phytase (EC 3.1.3.8 or 3.1.3.26); xylanase (EC 3.2.1.8); galactanase (EC 3.2.1.89); alpha-galactosidase (EC 3.2.1.22); protease (EC 3.4.-.-), phospholipase A1 (EC 3.1.1.32); phospholipase A2 (EC 3.1.1.4); lysophospholipase (EC 3.1.1.5); phospholipase C (3.1.4.3); phospholipase D (EC 3.1.4.4); and/or beta-glucanase (EC 3.2.1.4 or EC 3.2.1.6).In a particular embodiment these other enzymes are well-defined (as defined above for protease preparations).

Examples of antimicrobial peptides (AMP's) are CAP18, Leucocin A, Tritrpticin, Protegrin-1, Thanatin, Defensin, Lactoferrin, Lactoferricin, and Ovispirin such as Novispirin (Robert Lehrer, 2000), Plectasins, and Statins, including the compounds and polypeptides disclosed in WO 03/044049 and WO 03/048148, as well as variants or fragments of the above that retain antimicrobial activity.

Examples of antifungal polypeptides (AFP's) are the *Aspergillus giganteus*, and *Aspergillus niger* peptides, as well as variants and fragments thereof which retain antifungal activity, as disclosed in WO 94/01459 and WO 02/090384. Examples of polyunsaturated fatty acids are C18, C20 and C22 polyunsaturated fatty acids, such as arachidonic acid, docosohexaenoic acid, eicosapentaenoic acid and gamma-linoleic acid. Examples of reactive oxygen generating species are chemicals such as perborate, persulphate, or percarbonate; and enzymes such as an oxidase, an oxygenase or a syntethase.

Usally fat- and water-soluble vitamins, as well as trace minerals form part of a so-called premix intended for addition to the feed, whereas macro minerals are usually separately added to the feed. A premix enriched with a protease of the invention, is an example of an animal feed additive of the invention.

In a particular embodiment, the animal feed additive of the invention is intended for being included (or prescribed as having to be included) in animal diets or feed at levels of

0.01 to 10.0%; more particularly 0.05 to 5.0%; or 0.2 to 1.0% (% meaning g additive per 100 g feed). This is so in particular for premixes.

The following are non-exclusive lists of examples of these components:

Examples of fat-soluble vitamins are vitamin A, vitamin D3, vitamin E, and vitamin K, e.g. vitamin K3.

Examples of water-soluble vitamins are vitamin B12, biotin and choline, vitamin B1, vitamin B2, vitamin B6, niacin, folic acid and panthothenate, e.g. Ca-D-panthothenate.

Examples of trace minerals are manganese, zinc, iron, copper, iodine, selenium, and cobalt.

Examples of macro minerals are calcium, phosphorus and sodium.

The nutritional requirements of these components (exemplified with poultry and piglets/pigs) are listed in Table A of WO 01/58275. Nutritional requirement means that these components should be provided in the diet in the concentrations indicated.

In the alternative, the animal feed additive of the invention comprises at least one of the individual components specified in Table A of WO 01/58275. In the present context, at least one means either of, one or more of, one, or two, or three, or four and so forth up to all thirteen, or up to all fifteen individual components. More specifically, this at least one individual component is included in the additive of the invention in such an amount as to provide an in-feed-concentration within the range indicated in column four, or column five, or column six of Table A.

The above definition of "at least one", by the way, is generally valid all over the present patent application – of course on a by analogy basis, meaning that the upper limit of this definition, in the above example fifteen, of course should reflect the maximum number of choices given in each particular case.

The present invention also relates to animal feed compositions. Animal feed compositions or diets have a relatively high content of protein. Poultry and pig diets can be characterised as indicated in Table B of WO 01/58275, columns 2-3. Fish diets can be characterised as indicated in column 4 of this Table B. Furthermore such fish diets usually have a crude fat content of 200-310 g/kg.

An animal feed composition according to the invention has a crude protein content of 50-800 g/kg, and furthermore comprises at least one protease as claimed herein.

Furthermore, or in the alternative (to the crude protein content indicated above), the animal feed composition of the invention has a content of metabolisable energy of 10-30 MJ/kg; and/or a content of calcium of 0.1-200 g/kg; and/or a content of available phosphorus of 0.1-200 g/kg; and/or a content of methionine of 0.1-100 g/kg; and/or a content of methionine plus cysteine of 0.1-150 g/kg; and/or a content of lysine of 0.5-50 g/kg.

In particular embodiments, the content of metabolisable energy, crude protein, calcium, phosphorus, methionine, methionine plus cysteine, and/or lysine is within any one of ranges 2, 3, 4 or 5 in Table B of WO 01/58275 (R. 2-5).

Crude protein is calculated as nitrogen (N) multiplied by a factor 6.25, i.e. Crude protein (g/kg)= N (g/kg) x 6.25. The nitrogen content is determined by the Kjeldahl method (A.O.A.C., 1984, Official Methods of Analysis 14th ed., Association of Official Analytical Chemists, Washington DC).

Metabolisable energy can be calculated on the basis of the NRC publication Nutrient requirements in swine, ninth revised edition 1988, subcommittee on swine nutrition, committee on animal nutrition, board of agriculture, national research council. National Academy Press, Washington, D.C., pp. 2-6, and the European Table of Energy Values for Poultry Feed-stuffs, Spelderholt centre for poultry research and extension, 7361 DA Beekbergen, The Netherlands. Grafisch bedrijf Ponsen & looijen bv, Wageningen. ISBN 90-71463-12-5.

The dietary content of calcium, available phosphorus and amino acids in complete animal diets is calculated on the basis of feed tables such as Veevoedertabel 1997, gegevens over chemische samenstelling, verteerbaarheid en voederwaarde van voedermiddelen, Central Veevoederbureau, Runderweg 6, 8219 pk Lelystad. ISBN 90-72839-13-7.

In a particular embodiment, the animal feed composition of the invention contains at least one vegetable protein or protein source as defined above.

In still further particular embodiments, the animal feed composition of the invention contains 0-80% maize; and/or 0-80% sorghum; and/or 0-70% wheat; and/or 0-70% Barley; and/or 0-30% oats; and/or 0-40% soybean meal; and/or 0-10% fish meal; and/or 0-20% whey.

Animal diets can e.g. be manufactured as mash feed (non pelleted) or pelleted feed. Typically, the milled feed-stuffs are mixed and sufficient amounts of essential vitamins and minerals are added according to the specifications for the species in question. Enzymes can be added as solid or liquid enzyme formulations. For example, a solid enzyme formulation is typically added before or during the mixing step; and a liquid enzyme preparation is typically added after the pelleting step. The enzyme may also be incorporated in a feed additive or premix.

The final enzyme concentration in the diet is within the range of 0.01-200 mg enzyme protein per kg diet, for example in the range of 0.5-25, or 5-30, mg enzyme protein per kg animal diet.

The protease should of course be applied in an effective amount, i.e. in an amount adequate for improving solubilisation and/or improving nutritional value of feed. It is at

present contemplated that the enzyme is administered in one or more of the following amounts (dosage ranges): 0.01-200; 0.01-100; 0.05-100; 0.5-100; 1-100; 5-100; 10-100; 0.05-50; 1-50; or 0.10-10 – all these ranges being in mg protease enzyme protein per kg feed (ppm).

For determining mg enzyme protein per kg feed, the protease is purified from the feed composition, and the specific activity of the purified protease is determined using a relevant assay (see under protease activity, substrates, and assays). The protease activity of the feed composition as such is also determined using the same assay, and on the basis of these two determinations, the dosage in mg enzyme protein per kg feed is calculated.

The same principles apply for determining mg enzyme protein in feed additives. Of course, if a sample is available of the protease used for preparing the feed additive or the feed, the specific activity is determined from this sample (no need to purify the protease from the feed composition or the additive).

Detergent Compositions

The protease of the invention may be added to and thus become a component of a detergent composition.

The detergent composition of the invention may for example be formulated as a hand or machine laundry detergent composition including a laundry additive composition suitable for pre-treatment of stained fabrics and a rinse added fabric softener composition, or be formulated as a detergent composition for use in general household hard surface cleaning operations, or be formulated for hand or machine dishwashing operations.

In a specific aspect, the invention provides a detergent additive comprising the protease of the invention. The detergent additive as well as the detergent composition may comprise one or more other enzymes such as a lipase, a cutinase, an amylase, a carbohydrase, a cellulase, a pectinase, a mannanase, an arabinase, a galactanase, a xylanase, an oxidase, e.g., a laccase, and/or a peroxidase.

In general the properties of the chosen enzyme(s) should be compatible with the selected detergent, (i.e. pH-optimum, compatibility with other enzymatic and non-enzymatic ingredients, etc.), and the enzyme(s) should be present in effective amounts.

Suitable lipases include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Examples of useful lipases include lipases from *Humicola* (synonym *Thermomyces*), e.g. from *H. lanuginosa* (*T. lanuginosus*) as described in EP 258068 and EP 305216 or from *H. insolens* as described in WO 96/13580, a *Pseudomonas* lipase, e.g. from *P. alcaligenes* or *P. pseudoalcaligenes* (EP 218272), *P. cepacia* (EP 331376), *P. stutzeri* (GB 1,372,034), *P. fluorescens*, *Pseudomonas sp.* strain SD 705 (WO 95/06720 and WO 96/27002), *P. wisconsinensis* (WO 96/12012), a *Bacillus*

lipase, e.g. from *B. subtilis* (Dartois et al. (1993), Biochemica et Biophysica Acta, 1131, 253-360), *B. stearothermophilus* (JP 64/744992) or *B. pumilus* (WO 91/16422). Other examples are lipase variants such as those described in WO 92/05249, WO 94/01541, EP 407225, EP 260105, WO 95/35381, WO 96/00292, WO 95/30744, WO 94/25578, WO 95/14783, WO 95/22615, WO 97/04079 and WO 97/07202. Preferred commercially available lipase enzymes include Lipolase™ and Lipolase Ultra™ (Novozymes A/S).

Suitable amylases (alpha- and/or beta-) include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Amylases include, for example, alpha-amylases obtained from *Bacillus*, e.g. a special strain of *B. licheniformis*, described in more detail in GB 1,296,839. Examples of useful amylases are the variants described in WO 94/02597, WO 94/18314, WO 96/23873; and WO 97/43424, especially the variants with substitutions in one or more of the following positions: 15, 23, 105, 106, 124, 128, 133, 154, 156, 181, 188, 190, 197, 202, 208, 209, 243, 264, 304, 305, 391, 408, and 444. Commercially available amylases are Duramyl™, Termamyl™, Fungamyl™ and BAN™ (Novozymes A/S), Rapidase™ and Purastar™ (from Genencor International Inc.).

Suitable cellulases include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Suitable cellulases include cellulases from the genera *Bacillus*, *Pseudomonas*, *Humicola*, *Fusarium*, *Thielavia*, *Acremonium*, e.g. the fungal cellulases produced from *Humicola insolens*, *Myceliophthora thermophila* and *Fusarium oxysporum* disclosed in US 4,435,307, US 5,648,263, US 5,691,178, US 5,776,757 and WO 89/09259. Especially suitable cellulases are the alkaline or neutral cellulases having colour care benefits. Examples of such cellulases are cellulases described in EP 0 495257, EP 531372, WO 96/11262, WO 96/29397, WO 98/08940. Other examples are cellulase variants such as those described in WO 94/07998, EP 0 531 315, US 5,457,046, US 5,686,593, US 5,763,254, WO 95/24471, WO 98/12307 and WO 99/01544. Commercially available cellulases include Celluzyme™, and Carezyme™ (Novozymes A/S), Clazinase™, and Puradax HA™ (Genencor International Inc.), and KAC-500(B)™ (Kao Corporation).

Suitable peroxidases/oxidases include those of plant, bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Examples of useful peroxidases include peroxidases from *Coprinus*, e.g. from *C. cinereus*, and variants thereof as those described in WO 93/24618, WO 95/10602, and WO 98/15257. Commercially available peroxidases include Guardzyme™ (Novozymes).

The detergent enzyme(s) may be included in a detergent composition by adding separate additives containing one or more enzymes, or by adding a combined additive comprising all of these enzymes. A detergent additive of the invention, i.e. a separate additive or a combined additive, can be formulated e.g. as a granulate, a liquid, a slurry, etc.

Preferred detergent additive formulations are granulates, in particular non-dusting granulates, liquids, in particular stabilized liquids, or slurries.

Non-dusting granulates may be produced, e.g., as disclosed in US 4,106,991 and 4,661,452 and may optionally be coated by methods known in the art. Examples of waxy coating materials are poly(ethylene oxide) products (polyethyleneglycol, PEG) with mean molar weights of 1000 to 20000; ethoxylated nonylphenols having from 16 to 50 ethylene oxide units; ethoxylated fatty alcohols in which the alcohol contains from 12 to 20 carbon atoms and in which there are 15 to 80 ethylene oxide units; fatty alcohols; fatty acids; and mono- and di- and triglycerides of fatty acids. Examples of film-forming coating materials suitable for application by fluid bed techniques are given in GB 1483591. Liquid enzyme preparations may, for instance, be stabilized by adding a polyol such as propylene glycol, a sugar or sugar alcohol, lactic acid or boric acid according to established methods. Protected enzymes may be prepared according to the method disclosed in EP 238216.

The detergent composition of the invention may be in any convenient form, e.g., a bar, a tablet, a powder, a granule, a paste or a liquid. A liquid detergent may be aqueous, typically containing up to 70 % water and 0-30 % organic solvent, or non-aqueous.

The detergent composition comprises one or more surfactants, which may be non-ionic including semi-polar and/or anionic and/or cationic and/or zwitterionic. The surfactants are typically present at a level of from 0.1% to 60% by weight.

When included therein the detergent will usually contain from about 1% to about 40% of an anionic surfactant such as linear alkylbenzenesulfonate, alpha-olefinsulfonate, alkyl sulfate (fatty alcohol sulfate), alcohol ethoxysulfate, secondary alkanesulfonate, alpha-sulfo fatty acid methyl ester, alkyl- or alkenylsuccinic acid or soap.

When included therein the detergent will usually contain from about 0.2% to about 40% of a non-ionic surfactant such as alcohol ethoxylate, nonylphenol ethoxylate, alkylpolyglycoside, alkyldimethylamineoxide, ethoxylated fatty acid monoethanolamide, fatty acid monoethanolamide, polyhydroxy alkyl fatty acid amide, or N-acyl N-alkyl derivatives of glucosamine ("glucamides").

The detergent may contain 0-65 % of a detergent builder or complexing agent such as zeolite, diphosphate, triphosphate, phosphonate, carbonate, citrate, nitrilotriacetic acid, ethylenediaminetetraacetic acid, diethylenetriaminepentaacetic acid, alkyl- or alkenylsuccinic acid, soluble silicates or layered silicates (e.g. SKS-6 from Hoechst).

The detergent may comprise one or more polymers. Examples are carboxymethylcellulose, poly(vinylpyrrolidone), poly (ethylene glycol), poly(vinyl alcohol), poly(vinylpyridine-N-oxide), poly(vinylimidazole), polycarboxylates such as polyacrylates, maleic/acrylic acid copolymers and lauryl methacrylate/acrylic acid copolymers.

The detergent may contain a bleaching system which may comprise a H_2O_2 source such as perborate or percarbonate which may be combined with a peracid-forming bleach activator such as tetraacetylethylenediamine or nonanoyloxybenzenesulfonate. Alternatively, the bleaching system may comprise peroxyacids of e.g. the amide, imide, or sulfone type.

The enzyme(s) of the detergent composition of the invention may be stabilized using conventional stabilizing agents, e.g., a polyol such as propylene glycol or glycerol, a sugar or sugar alcohol, lactic acid, boric acid, or a boric acid derivative, e.g., an aromatic borate ester, or a phenyl boronic acid derivative such as 4-formylphenyl boronic acid, and the composition may be formulated as described in e.g. WO 92/19709 and WO 92/19708.

The detergent may also contain other conventional detergent ingredients such as e.g. fabric conditioners including clays, foam boosters, suds suppressors, anti-corrosion agents, soil-suspending agents, anti-soil redeposition agents, dyes, bactericides, optical brighteners, hydrotropes, tarnish inhibitors, or perfumes.

It is at present contemplated that in the detergent compositions any enzyme, in particular the enzyme of the invention, may be added in an amount corresponding to 0.01-100 mg of enzyme protein per liter of wash liqour, preferably 0.05-5 mg of enzyme protein per liter of wash liqour, in particular 0.1-1 mg of enzyme protein per liter of wash liqour.

The enzyme of the invention may additionally be incorporated in the detergent formulations disclosed in WO 97/07202.

Various Embodiments

In a particular embodiment, the polypeptide of the invention does not encompass (i.e., excludes): a) amino acids -186 to 188, -167 to 188, or 1-188 of SEQ ID NO: 4; b) amino acids 1-188 of SEQ ID NO: 26; c) the protease from *Nocardiopsis dassonvillei* NRRL 18133 which is described in WO 88/03947, preferably having a Molecular Weight (MW) by SDS-PAGE of 20,500 Dalton, and isoelectric points, pI, of 9.15 and 8.2; d) the protease from *Nocardiopsis sp.* which is described in JP 2255081 A, preferably having a MW by SDS electrophoresis of 21,000 Da and an optimum pH of 10-12; and/or e) the protease derived from the strain ZIMET 43647 of the species *Nocardiopsis dassonvillei* described by GDR patent no. DD 2,004,328.

In another particular embodiment, the nucleic acid sequence of the invention, the nucleic acid sequence of the invention does not encompass (i.e., excludes): a) nucleotides 1-1122 and/or 559-1122 of SEQ ID NO: 3; and/or b) nucleotides 1-1596 and/or 900-1466 of SEQ ID NO: 25.

The invention also relates to each of the embodiments listed below. All what is said hereinbefore, in particular as regards the *Brachysporiella* protease, SEQ ID NO: 2, and its encoding nucleotide sequence SEQ ID NO: 1, applies by analogy to each of the below listed

embodiments. This includes, e.g., the procedures for aligning amino acid sequences, the numbering rules by reference to the *Gliocladium* protease sequence of Fig. 1, determination of degree of identity, hybridization conditions, as well as various particular embodiments reciting different percentage identities, different hybridization conditions, etc. Specifically included herein are also, e.g., claims relating to isolated nucleic acid sequences encoding the below polypeptides, claims to the corresponding nucleic acid constructs, the corresponding recombinant host cells, corresponding methods for producing the proteases, corresponding transgenic plants and animals, corresponding use in animal feed, corresponding animal feed additives, corresponding animal feed compositions, and corresponding detergent compositions. The term "corresponding" in this context refers to the claims as filed herewith, the polypeptide or nucleic acid sequence part thereof however being replaced by the below definitions of the polypeptides, respectively the nucleic acid sequences, of the invention.

An isolated polypeptide having protease activity and comprising an amino acid sequence which has a percentage of identity to SEQ ID NO: 2 of at least 62%.

An isolated polypeptide having protease activity and comprising an amino acid sequence which, when aligned according to Fig. 1, comprises at least one of the following:
CDEFGHIKLMNPQRSTVWY61; ACDEFGHIKLMNPQRSTWY63;
ACDEFHIKLMNPQSTVWY77; ACDEFGHKLMNPQRSTVWY111;
ACDEFGHILMNPQSTVWY114; ACDEFGHIKLMNQRSTVWY121;
ACDEFGHIKLMNPQSTVWY130; ACDEFGHIKLMNPQSTVWY162;
ACDEFGHIKLMNPQRSVWY163; ACDEFGHIKLMNQSTVWY174; and/or
ACDEFGHIKLMPRSTVWY177;
wherein the numbering of each amino acid residue corresponds to the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6).

An isolated polypeptide having protease activity and comprising an amino acid sequence which, when aligned according to Fig. 1, comprises at least one of the following:
ACDEFGHILMNPQSTVWY114; ACDEFGHIKLMNQRSTVWY121;
ACDEFGHIKLMNPQSTVWY130; ACDEFGHIKLMNPQSTVWY162;
ACDEFGHIKLMNPQRSVWY163; ACDEFGHIKLMNQSTVWY174; and/or
ACDEFGHIKLMPRSTVWY177,
wherein the numbering of each amino acid residue corresponds to the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6).

An isolated polypeptide having protease activity and comprising an amino acid sequence which, when aligned according to Fig. 1, comprises at least one of the following:
S61; I63; N77 or T77; (V111 or L111); (L114 or Y114); (S121 or D121); (Q130 or M130); (N162 or T162 or S162); (S163 or R163); E174; and/or (S177 or E177);

wherein the numbering of each amino acid residue corresponds to the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6).

An isolated polypeptide having protease activity and comprising an amino acid sequence which, when aligned according to Fig. 1, comprises at least one of the following: (L114 or Y114); (S121 or D121); (Q130 or M130); (N162 or T162 or S162); (S163 or R163); E174; and/or (S177 or E177) wherein the numbering of each amino acid residue corresponds to the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6).

An isolated polypeptide having protease activity and comprising an amino acid sequence which, when aligned according to Fig. 1, comprises at least one of the following: (S61 + I63 + (N77 or T77) + (V111 or L111) + (L114 or Y114) + (S121 or D121) + (Q130 or M130) + (N162 or T162 or S162) + (S163 or R163) + E174 + (S177 or E177)), wherein the numbering of each amino acid residue corresponds to the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6).

An isolated polypeptide having protease activity and comprising an amino acid sequence which, when aligned according to Fig. 1, comprises at least one of the following: ((L114 or Y114) + (S121 or D121) + (Q130 or M130) + (N162 or T162 or S162) + (S163 or R163) + E174 + (S177 or E177)), wherein the numbering of each amino acid residue corresponds to the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6).

An isolated polypeptide having protease activity and comprising an amino acid sequence which, when aligned according to Fig. 1, comprises at least one of the following: (N162 or T162); and/or S163)), wherein the numbering of each amino acid residue corresponds to the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6).

An isolated polypeptide having protease activity and having an amino acid sequence which comprises amino acids 1-186 of SEQ ID NO: 2, amino acids 1-186 of SEQ ID NO: 6, amino acids 1-186 of SEQ ID NO: 8, amino acids 1-185 of SEQ ID NO: 10, or amino acids 1-65 of SEQ ID NO: 12; or variants or fragments thereof.

An isolated polypeptide having protease activity, selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence of at least 55% identity to amino acids 1-185 of SEQ ID NO: 10;

(b) a polypeptide which is encoded by a nucleic acid sequence which hybridizes under low stringency conditions with

(i) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 16128,

(ii) nucleotides 652-1206, 139-1206, or 82-1206 of SEQ ID NO: 9,

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, or

(iv) a complementary strand of (i), (ii) or (iii);

(c) a variant of the polypeptide having an amino acid sequence of amino acids 1 to 185, -171 to 185, or -190 to 185 of SEQ ID NO: 10 comprising a substitution, deletion, and/or insertion of one or more amino acids;

(d) an allelic variant of (a) or (b); and

(e) a fragment of (a), (b), or (d) that has protease activity.

An isolated polypeptide having protease activity, selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence of at least 54% identity to amino acids 1-186 of SEQ ID NO: 8;

(b) a polypeptide which is encoded by a nucleic acid sequence which hybridizes under low stringency conditions with

(i) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 16052,

(ii) nucleotides 638-1195, 134-1195, or 77-1195 of SEQ ID NO: 7,

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, or

(iv) a complementary strand of (i), (ii) or (iii);

(c) a variant of the polypeptide having an amino acid sequence of amino acids 1 to 186, -168 to 186, or -187 to 186 of SEQ ID NO: 8 comprising a substitution, deletion, and/or insertion of one or more amino acids;

(d) an allelic variant of (a) or (b); and

(e) a fragment of (a), (b), or (d) that has protease activity.

An isolated polypeptide having protease activity, selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence of at least 52% identity to amino acids 1-186 of SEQ ID NO: 6;

(b) a polypeptide which is encoded by a nucleic acid sequence which hybridizes under low stringency conditions with

(i) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 16051,

(ii) nucleotides 636-1193, 114-1193, or 57-1193 of SEQ ID NO: 5,

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, or

(iv) a complementary strand of (i), (ii) or (iii);

(c) a variant of the polypeptide having an amino acid sequence of amino acids 1 to 186, -174 to 186, or -193 to 186 of SEQ ID NO: 6 comprising a substitution, deletion, and/or insertion of one or more amino acids;

(d) an allelic variant of (a) or (b); and

(e) a fragment of (a), (b), or (d) that has protease activity.

An isolated polypeptide having protease activity, selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence of at least 48% identity to amino acids 1-186 of SEQ ID NO: 2;

(b) a polypeptide which is encoded by a nucleic acid sequence which hybridizes under low stringency conditions with

(i) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 15509,

(ii) nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1,

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, or

(iv) a complementary strand of (i), (ii) or (iii);

(c) a variant of the polypeptide having an amino acid sequence of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2 comprising a substitution, deletion, and/or insertion of one or more amino acids;

(d) an allelic variant of (a) or (b); and

(e) a fragment of (a), (b), or (d) that has protease activity.

An isolated polypeptide having protease activity, selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence of at least 47% identity to amino acids 1-65 of SEQ ID NO: 12;

(b) a polypeptide which is encoded by a nucleic acid sequence which hybridizes under low stringency conditions with

(i) nucleotides 355-549, or 1-549 of SEQ ID NO: 11,

(ii) a subsequence of (i) or (ii) of at least 100 nucleotides, or

(iii) a complementary strand of (i), (ii) or (iii);

(c) a variant of the polypeptide having an amino acid sequence of amino acids 1 to 65, or -118 to 65 of SEQ ID NO: 12 comprising a substitution, deletion, and/or insertion of one or more amino acids;

(d) an allelic variant of (a) or (b); and

(e) a fragment of (a), (b), or (d) that has protease activity.

An isolated nucleic acid sequence comprising a nucleic acid sequence which encodes a polypeptide having protease activity, and which

(a) hybridizes under low stringency conditions with

(i) nucleotides 652-1206, 139-1206, or 82-1206 of SEQ ID NO: 9,

(ii) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 16128,

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, and/or

(iv) a complementary strand of (i), (ii), or (iii); and/or

(b) has a degree of identity to nucleotides 652-1206, 139-1206, or 82-1206 of SEQ ID NO: 9, of at least 64%.

An isolated nucleic acid sequence comprising a nucleic acid sequence which encodes a polypeptide having protease activity, and which

(a) hybridizes under low stringency conditions with

(i) nucleotides 638-1195, 134-1195, or 77-1195 of SEQ ID NO: 7,

(ii) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 16052,

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, and/or

(iv) a complementary strand of (i), (ii), or (iii); and/or

(b) has a degree of identity to nucleotides 638-1195, 134-1195, or 77-1195 of SEQ ID NO: 7, of at least 64%.

An isolated nucleic acid sequence comprising a nucleic acid sequence which encodes a polypeptide having protease activity, and which

(a) hybridizes under low stringency conditions with

(i) nucleotides 636-1193, 114-1193, or 57-1193 of SEQ ID NO: 5,

(ii) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 16051,

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, and/or

(iv) a complementary strand of (i), (ii), or (iii); and/or

(b) has a degree of identity to nucleotides 636-1193, 114-1193, or 57-1193 of SEQ ID NO: 5, of at least 65%.

An isolated nucleic acid sequence comprising a nucleic acid sequence which encodes a polypeptide having protease activity, and which

(a) hybridizes under low stringency conditions with

(i) nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1,

(ii) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 15509,

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, and/or

(iv) a complementary strand of (i), (ii), or (iii); and/or

(b) has a degree of identity to nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1, of at least 61%.

An isolated nucleic acid sequence comprising a nucleic acid sequence which encodes a polypeptide having protease activity, and which

(a) hybridizes under low stringency conditions with

(i) nucleotides 355-549, or 1-549 of SEQ ID NO: 11,

(ii) a subsequence of (i) or (ii) of at least 100 nucleotides, and/or

(iii) a complementary strand of (i), (ii), or (iii); and/or

(b) has a degree of identity to nucleotides 652355-549, or 1-549 of SEQ ID NO: 11, of at least 57%.

An isolated polypeptide having protease activity, selected from the group consisting of:

(a) a polypeptide having an amino acid sequence which has a degree of identity to amino acids 1 to 186 of SEQ ID NO: 2 of at least 65%;

(b) a polypeptide which is encoded by a nucleic acid sequence which hybridizes under low stringency conditions with

(i) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 15509,

(ii) nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1,

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, or

(iv) a complementary strand of (i), (ii) or (iii);

(c) a variant of the polypeptide having an amino acid sequence of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2 comprising a substitution, deletion, and/or insertion of one or more amino acids;

(d) an allelic variant of (a) or (b); and

(e) a fragment of (a), (b), or (d) that has protease activity.

An isolated nucleic acid sequence comprising a nucleic acid sequence which

(a) encodes the polypeptide of the above embodiment;

(b) encodes a polypeptide having protease activity, and which hybridizes under low stringency conditions with

(i) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 15509,

(ii) nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1,

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, and/or

(iv) a complementary strand of (i), (ii), or (iii);

(c) encodes a polypeptide having protease activity and which has a degree of identity to nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1 of at least 70%.

An isolated nucleic acid sequence produced by

(a) hybridizing a DNA under low stringency conditions with

(i) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 15509,

(ii) nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1;

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, or

(iv) a complementary strand of (i), (ii) or (iii); and

(b) isolating the nucleic acid sequence.

A nucleic acid construct comprising the nucleic acid sequence of either of the above two embodiments, operably linked to one or more control sequences that direct the production of the polypeptide in a suitable expression host.

A recombinant expression vector comprising the abovenucleic acid construct.

A recombinant host cell comprising the above nucleic acid construct or the above vector.

A method for producing the polypeptide of the above embodiment, the method comprising (a) cultivating the above recombinant host cell to produce a supernatant comprising the polypeptide; and (b) recovering the polypeptide.

A transgenic plant, or plant part, capable of expressing the protease of the above embodiment.

A transgenic, non-human animal, or products, or elements thereof, being capable of expressing the protease of the above embodiment.

Deposit of Biological Material

The following biological materials have been deposited under the terms of the Budapest Treaty with the CGMCC (China General Microbiological Culture Collection Center, Institute of Microbiology, Beijing 100080, China), DSM (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, 38124 Braunschweig, Germany), and CBS (Centraal-bureau voor Schimmelcultures, Uppsalalaan 8, NL-3584 Utrecht, the Netherlands) respectively, and given the following accession numbers:

Deposit	Accession Number	Date of Deposit
Brachysporiella sp.	CGMCC 0865	December 19, 2002
Escherichia coli	DSM 15509	March 18, 2003
Gliocladium sp.	CBS 114001	November 17, 2003
Escherichia coli	DSM 16051	November 24, 2003
Periconia sp.	CBS 114002	November 17, 2003
Escherichia coli	DSM 16052	November 24, 2003
Periconia sp.	CBS 114000	November 17, 2003

Curvularia sp.	CBS 114003	November 17, 2003
Escherichia coli	DSM 16128	January 8, 2004

The strains have been deposited under conditions that assure that access to the cultures will be available during the pendency of this patent application to one determined by the Commissioner of Patents and Trademarks to be entitled thereto under 37 C.F.R. §1.14 and 35 U.S.C. §122. The deposits represent substantially pure cultures of the deposited strains. The deposits are available as required by foreign patent laws in countries wherein counterparts of the subject application or its progeny are filed. However, it should be understood that the availability of a deposit does not constitute a license to practice the subject invention in derogation of patent rights granted by governmental action.

The *Brachysporiella* strain CGMCC 0865 was isolated from dead branches of an unidentified plant in China in October 1998. The strains *Gliocladium sp.* CBS 114001, *Periconia sp.* CBS 114002, *Periconia sp.* CBS 114000, and *Curvularia sp.* CBS 114003 were isolated from dead wood in China in 1998 and 1999.

The invention described and claimed herein is not to be limited in scope by the specific embodiments herein disclosed, since these embodiments are intended as illustrations of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. In the case of conflict, the present disclosure including definitions will control.

Various references are cited herein, the disclosures of which are incorporated by reference in their entireties.

Examples

Example 1: Characterization of a protease derived from *Brachysporiella*

The gene of SEQ ID NO: 1 encoding the protease of SEQ ID NO: 2 was isolated from *Brachysporiella gayana* CGMCC 0865, expressed in *Aspergillus oryzae*, and the protease was purified from the culture broth. The purified mature protease is an α-lytic protease like enzyme (peptidase family S2A/SE1), M_r = 23kDa (SDS-PAGE), GPMAW: 18345Da, inhibited by PMSF. (GPMAW = General Protein/Mass Analysis for Windows; a software package from Lighthouse data, Odense, Denmark).

Protease assays

1) pNA Assay:

pNA substrate: Suc-AAPF-pNA (Bachem L-1400)

Temperature: room temperature (20°C)

Assay buffers: 100mM succinic acid, 100mM HEPES, 100mM CHES, 100mM CABS, 1mM $CaCl_2$, 150mM KCl, 0.01% Triton X-100 adjusted to pH-values 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0, and 12.0 with HCl or NaOH.

20µl protease (diluted in 0.01% Triton X-100) was mixed with 100µl assay buffer. The assay was started by adding 100µl pNA substrate (50mg dissolved in 1.0ml DMSO and further diluted 45x with 0.01% Triton X-100). The increase in OD_{405} was monitored as a measure of the protease activity ($mOD_{405} = 10^{-3}$ OD_{405}).

2) Protazyme AK assay:

Substrate: Protazyme AK tablet (from Megazyme).

Temperature: controlled (assay temperature)

Assay buffers: 100mM succinic acid, 100mM HEPES, 100mM CHES, 100mM CABS, 1mM $CaCl_2$, 150mM KCl, 0.01% Triton X-100 adjusted to pH-values 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0 and 11.0 with HCl or NaOH.

A Protazyme AK tablet (from Megazyme) was suspended in 2.0ml 0.01% Triton X-100 by gentle stirring. 500µl of this suspension and 500µl assay buffer were mixed in an Eppendorf tube and placed on ice. 20µl protease sample (diluted in 0.01% Triton X-100) was added. The assay was initiated by transferring the Eppendorf tube to an Eppendorf thermomixer, which was set to the assay temperature. The tube was incubated for 15 minutes on the Eppendorf thermomixer at its highest shaking rate. The incubation was stopped by transferring the tube back to the ice bath. Then the tube was centrifuged in an icecold centrifuge for a few minutes and 200µl supernatant was transferred to a microtiter plate. OD_{650} was read as a measure of protease activity. A buffer blind was included in the assay (instead of enzyme).

pH Activity profile, pH-stability, temperature-activity profile

The pNA assay was used for obtaining the pH-activity profile (Table 1) as well as the pH-stability profile (Table 2). For the pH-stability profile the protease was diluted 10x in the assay buffers and incubated for 2 hours at 37°C. After incubation the protease samples were transferred to the same pH, before assay for residual activity, by dilution in one of the assay buffers. The Protazyme AK (cross-linked and dyed casein) assay was used for obtaining the temperature-activity profile at pH 9 (Table 3).

Table 1: pH-profile

pH	mOD_{405}/min	Relative Activity
2.0	1.000	0.002

pH	mOD_{405}/min	Relative Activity
2.5	0.175	0.000
3.0	0.264	0.001
3.5	0.713	0.001
4.0	1.750	0.003
5.0	14.42	0.028
6.0	98.01	0.189
7.0	291.99	0.562
8.0	399.71	0.770
9.0	429.05	0.826
10.0	501.36	0.966
11.0	519.17	1.000

Table 2: pH-stability

pH (after 2 hours at 37°C)	mOD/min	Residual activity
2.0	13.54	0.03
2.5	10.68	0.02
3.0	126.48	0.26
3.5	374.00	0.76
4.0	462.82	0.94
5.0	491.52	1.00
6.0	476.11	0.97
7.0	500.43	1.01
8.0	494.26	1.00
9.0	489.76	0.99
10.0	485.14	0.98
11.0	440.71	0.89
12.0	7.81	0.02
9.0 (after 2 hours at 5°C)	493.29	1.00

Table 3: Temperature-profile (pH 9.0)

Temperature	Enzyme Activity (OD_{650}, sample minus blind)	Relative Activity
15	0.095	0.13

Temperature	Enzyme Activity (OD_{650}, sample minus blind)	Relative Activity
25	0.154	0.21
37	0.263	0.35
50	0.522	0.70
60	0.746	1.00
70	0.169	0.23

Example 2: Performance of the protease derived from *Brachysporiella* in a monogastric *in vitro* digestion model

The performance of a purified preparation of the *Brachysporiella* protease having SEQ ID NO: 2 (prepared as described in Example 1) was tested in an *in vitro* model simulating the digestion in monogastric animals. In particular, the protease was tested for its ability to improve solubilisation and digestion of maize/-SBM (maize/-Soybean meal) proteins.

The *in vitro* system consisted of 20 flasks in which maize/-SBM substrate was initially incubated with HCl/pepsin - simulating gastric digestion - and subsequently with pancreatin – simulating intestinal digestion. 15 of the flasks were dosed with the protease at the start of the gastric phase whereas the remaining flasks served as blanks. At the end of the intestinal incubation phase samples of *in vitro* digesta were removed and analysed for solubilised and digested protein.

Outline of *in vitro* digestion procedure

Components added	pH	Temperature	Time course	Simulated digestion phase
10 g maize/-SBM substrate (6:4), 41 ml HCl (0.105M)	3.0	40°C	t=0 min	Mixing
5 ml HCl (0.105M) / pepsin (3000 U/g substrate), 1 ml *Brachysporiella* protease (various dosages of mg protease enzyme protein/kg diet)	3.0	40°C	t=30 min	Gastric digestion
16 ml H_2O	3.0	40°C	t= 1 hour	Gastric digestion
7 ml NaOH (0,39M)	6.8	40°C	t=1.5 hours	Intestinal digestion
5 ml $NaHCO_3$ (1M) / pancreatin (8 mg /g diet)	6.8	40°C	t=2 hours	Intestinal digestion

Components added	pH	Temperature	Time course	Simulated digestion phase
Terminate incubation	7.0	40°C	t=6 hours	

Conditions

Substrate: 4 g SBM, 6 g maize (premixed)

pH: 3.0 stomach step/ 6.8-7.0 intestinal step

HCl: 0.105 M for 1.5 hours (i.e. 30 min HCl-substrate premixing)

pepsin: 3000 U /g diet for 1 hour

pancreatin: 8 mg/g diet for 4 hours

temperature: 40°C.

Replicates: 5

Solutions

0.39 M NaOH

0.105 M HCl

0.105 M HCl containing 6000 U pepsin per 5 ml

1 M $NaHCO_3$ containing 16 mg pancreatin per ml

125 mM NaAc-buffer, pH 6.0

Enzyme protein determinations

The amount of protease enzyme protein is calculated on the basis of the A_{280} values and the amino acid sequences (amino acid compositions) using the principles outlined in S.C. Gill & P.H. von Hippel, Analytical Biochemistry 182, 319-326, (1989).

Experimental procedure for *in vitro* model

The experimental procedure was according to the above outline. pH was measured at time 1, 2.5, and 5.5 hours. Incubations were terminated after 6 hours and samples of 30 ml were removed and placed on ice before centrifugation (10000 x g, 10 min, 4°C). Supernatants were removed and stored at -20°C.

Analysis

All samples were analysed for % Degree of Hydrolysis (DH) with the OPA method as well as content of solubilised and digested protein using gelfiltration.

DH determination by the OPA-method

The Degree of Hydrolysis (DH) of protein in different samples was determined using an semi-automated microtiter plate based colorimetric method (Nielsen,P.M.; Petersen,D.; Dambmann,C. Improved method for determining food protein degree of hydrolysis. J.Food Sci. 2001, 66, 642-646). The OPA reagent was prepared as follows: 7.620 g di-Na tetraborate decahydrate and 200 mg sodiumdodecyl sulphate (SDS) were dissolved in 150 ml deionised water. The reagents were completely dissolved before continuing. 160 mg o-phthal-dialdehyde 97% (OPA) was dissolved in 4 ml ethanol. The OPA solution was

transferred quantitatively to the above-mentioned solution by rinsing with deionised water. 176 mg dithiothreitol 99% (DTT) was added to the solution that was made up to 200 ml with deionised water. A serine standard (0.9516 meqv/l) was prepared by solubilising 50 mg serine (Merck, Germany) in 500 ml deionized water.

The sample solution was prepared by diluting each sample to an absorbance (280 nm) of about 0.5. Generally, supernatants were diluted (100 ×) using an automated Tecan dilution station (Männedorf, Switzerland). All other spectrophotometer readings were performed at 340 nm using deionized water as the control. 25 µl of sample, standard and blind was dispensed into a microtiter plate. The microtiter plate was inserted into an iEMS MF reader (Labsystems, Finland) and 200µl of OPA reagent was automatically dispensed. Plates were shaken (2 min; 700 rpm) before measuring absorbance. Finally, the DH was calculated. Eightfold determination of all samples was carried out.

Estimation of solubilised and digested protein

The content of solubilised protein in supernatants from *in vitro* digested samples was estimated by quantifying crude protein (CP) using gel filtration HPLC. Supernatants were thawed, filtered through 0.45 µm polycarbonate filters and diluted (1:50, v/v) with H_2O. Diluted samples were chromatographed by HPLC using a Superdex Peptide PE (7.5 x 300 mm) gelfiltration column (Global). The eluent used for isocratic elution was 50 mM sodium phosphate buffer (pH 7.0) containing 150 mM NaCl. The total volume of eluent per run was 26 ml and the flow rate was 0.4 ml/min. Elution profiles were recorded at 214 nm and the total area under the profiles was determined by integration. Elution profiles were recorded at 214 nm and the total area under the profiles was determined by integration. To estimate protein content from integrated areas, a calibration curve (R^2=0.9993) was made from a dilution series of an *in vitro* digested reference maize/-SBM sample with known total protein content. The protein determination in this reference sample was carried out by the Kjeldahl method (determination of % nitrogen; A.O.A.C. (1984) Official Methods of Analysis 14th ed., Washington DC).

The content of digested protein was estimated by integrating the chromatogram area corresponding to peptides and amino acids having a molecular mass of 1500 dalton or below (Savoie, L.; Gauthier, S.F. Dialysis Cell For The In-vitro Measurement Of Protein Digestibility. J. Food Sci. 1986, 51, 494-498; Babinszky, L.; Van, D.M.J.M.; Boer, H.; Den, H.L.A. An In-vitro Method for Prediction of The Digestible Crude Protein Content In Pig Feeds. J. Sci. Food Agr. 1990, 50, 173-178; Boisen, S.; Eggum, B.O. Critical Evaluation of In-vitro Methods for Estimating Digestibility in Simple-Stomach Animals. Nutrition Research Reviews 1991, 4, 141-162). To determine the 1500 dalton dividing line, the gel fitration column was calibrated using cytochrome C, aprotinin, gastrin I, and substance P as molecular mass standards.

Results

The results shown in Tables 1 and 2 below indicate that the protease increased digestible protein significantly in the doses 25, 50 and 100 mg Enzyme Protein (EP) per kg feed. Soluble protein was increased significantly at 50 and 100 mg EP/kg feed and numerically increased at 25 mg EP/kg feed. Degree of Hydrolysis (DH) was increased significantly in the dosage 100 mg EP/kg and numerically increased at 25 and 50 mg EP/kg feed.

Table 1

Degree of Hydrolysis (DH), absolute and relative values

		Of total protein			Relative to blank		
Sample	n	%DH		SD	%DH		%CV
Blank	4	26.85	a	0.38	100.00	a	1.42
25 mg EP	5	27.47	a	0.42	102.34	a	1.52
50 mg EP	5	28.25	ab	0.30	105.23	ab	1.06
100 mg EP	5	28.62	b	0.44	106.60	b	1.55

Different letters within the same column indicate significant differences (1-way ANOVA, Tukey-Kramer test, P<0.05). SD = Standard Deviation. %CV = Coefficient of Variance = (SD/mean value) x 100%

Table 2

Solubilised and digested crude protein measured by ÄKTA HPLC, absolute and relative values.

		Of total protein						Relative to blank					
Sample	n	% dig. CP		SD	% sol. CP		SD	% dig. CP		CV%	% sol. CP		CV%
Blank	9	62.7	a	1.3	90.5	a	1.1	100.0	a	2.1	100.0	a	1.2
25 mg EP	5	65.3	b	0.6	92.1	ab	1.0	104.1	b	0.9	101.7	ab	1.1
50 mg EP	5	66.6	b	1.2	94.1	bc	1.6	106.3	b	1.8	104.0	bc	1.7
100 mg EP	5	67.0	b	1.1	94.1	bc	1.7	106.9	b	1.7	104.0	bc	1.9

Different letters within the same column indicate significant differences (1-way ANOVA, Tukey-Kramer test, P<0.05). SD = Standard Deviation. %CV = Coefficient of Variance = (SD/mean value) x 100%

Example 3: Isolation of a related protease from *Gliocladium*

Fungal proteases related to the *Brachysporiella* protease (SEQ ID NO: 2) have

proven difficult to find, as the yield is low, and because they are usually found in admixture with other types of proteases in cultures of wild type strains. Therefore, molecular screening was used in the search for related fungal proteases.

The following probes were designed for PCR screening based on alignments of SEQ ID NO: 1 and SEQ ID NO: 3, as well as the corresponding amino acid sequences SEQ ID NO: 2 and SEQ ID NO: 4:

10R0sa: 5' gcn atg aa(a/g) cgn ga(t/c)(c/t)t 3' (SEQ ID NO: 13)

10Rd3R: 5' gt(gtc) g(at)(gt) cc(gt) ga(cg) cgg ca(cg) a 3' (SEQ ID NO: 14)

4-6 agar plugs with a fully grown culture of Gliocladium sp. strain CBS 114001 were used to inoculate a 500 ml shake flask with 50 ml of the FG-4 medium (30 g Soymeal, 15 g Maltose, 5 g Peptone, 1000 ml H_2O, 1 g olive oil). The flask was incubated at 25°C and 160 rpm for 72 hours. The mycelium was harvested by centrifugation at 4°C, and quickly frozen in liquid N_2 and stored at -80°C.

RNA was isolated by using RNeasy Mini Kit (Qiagen company, Cat. No. 74904). The corresponding cDNA was synthesized using the 3' Rapid Amplification of cDNA End System (3' RACE) (Life Technologies, Cat. No. 18373-027).

The first PCR was performed by using 10R0sa (SEQ ID NO: 13) and AUAP (Abridged Universal Amplification Primer: 5' ggc cac gcg tcg act agt ac 3' (SEQ ID NO: 23), provided with the 3' RACE kit) as primer pair and cDNA as template. The nested PCR was performed by using 10R0sa (SEQ ID NO: 13) and 10Rd3R (SEQ ID NO: 14) and the first PCR solution as template. The PCR program was: 94°C for 3min; 35 cycles of 94°C for 45s, 55°C for 30s, 72°C for 1min; final extension at 72°C for 10min. An alternative primer similar to 10R0sa is: 5' gcn atg aa(a/g) cgn ga(t/c) ctn 3'.

A sequence of about 750 base pairs was amplified and recovered and confirmed by DNA sequencing to be related to SEQ ID NO: 1. Based on this, the following new primers were designed for cloning of the 5' and the 3'end:

733ras1: 5' act gtg gcg aaa gtc tca 3' (SEQ ID NO: 15)

733ras2: 5' cgc act tca aac tca ga 3' (SEQ ID NO: 16)

733ras3: 5' gtt gtc aag tgc cac ctt 3' (SEQ ID NO: 17)

733rf1: 5' tcg ttg ctc tat tgg ctt 3' (SEQ ID NO: 18)

733rf2: 5' ctt ctc tgg ttc cgt ctt 3' (SEQ ID NO: 19)

For the 3' end cloning, PCR was performed by using primer pair 733rf1 (SEQ ID NO: 18) with AUAP (SEQ ID NO: 23), and the first PCR solution as a template, viz. (10R0sa (SEQ ID NO: 13) and AUAP (SEQ ID NO: 23) as primer pair and cDNA as template). A specific amplification of 750 base pairs was obtained. The fragment was identified as related to SEQ ID NO: 1.

The 3' primers were thus designed for full length cloning:

733ras00: 5' act aac cat tgc cca ta 3' (SEQ ID NO: 20)

733ras01: 5' ctt agg tac gtc aac cat 3' (SEQ ID NO: 21)

For 5'end cloning, cDNA was synthesized with initiation of 733ras1 (SEQ ID NO: 15). by using the 5' Rapid Amplification of cDNA End System (5' RACE) (Life Technologies, Cat. No. 18374-041). The primary PCR was performed using cDNA as template and with primer pair 733ras1 (SEQ ID NO: 15) and AAP (Abridged Anchor Primer: 5' ggc cac gcg tcg act agt acg ggi igg gii ggg iig 3' (SEQ ID NO: 24), supplied by the 5'RACE system). Nested PCR was performed by using 733ras3 (SEQ ID NO: 17) and AUAP (SEQ ID NO: 23) as primer pair and the primary PCR as template. A specific amplification was obtained at 500 base pairs. It was identified as a sequence related to SEQ ID NO: 1, and then the 5' primer was designed for full length cloning:

733rs00: 5' agc atc tac aac atc gac a 3' (SEQ ID NO: 22)

Finally, the full length gene was cloned by PCR with 733rs00 (SEQ ID NO: 22) and 733ras00 (SEQ ID NO: 20). The PCR program was: 94°C, 3mins; 35 cycles of 94°C, 1min; 55°C, 30secs; 72°C, 1.5min; final extension at 72°C, 10mins. A specific fragment at ~1.3kb was amplified, sequenced, and identified as related to SEQ ID NO: 1.

The sequence obtained (SEQ ID NO: 5) had a coding sequence of 1137 base pairs. The corresponding deduced amino acid sequence is SEQ ID NO: 6.

Example 4: Isolation of a related protease from *Periconia* CBS 114002

RNA was isolated from *Periconia sp.* strain CBS 114002, and a gene encoding a protease related to SEQ ID NO: 1 isolated, as generally described in Example 3.

The sequence obtained (SEQ ID NO: 7) had a coding sequence of 1137 base pairs. The corresponding deduced amino acid sequence is SEQ ID NO: 8, the predicted signal peptide of which is amino acids 1-23, and the predicted mature peptide of which is amino acids 188-373.

Example 5: Isolation of a related protease from *Curvularia*

RNA was isolated from *Curvularia lunata* strain CBS 114003, and a gene encoding a protease related to SEQ ID NO: 1 isolated, as generally described in Example 3.

The sequence obtained (SEQ ID NO: 9) had a coding sequence of 1259 base pairs. The corresponding deduced amino acid sequence is SEQ ID NO: 10.

Example 6: Isolation of a related protease from *Periconia* CBS 114000

RNA was isolated from *Periconia sp.* strain CBS 114000, and a partial DNA fragment related to SEQ ID NO: 1 isolated, as generally described in Example 3.

The first PCR reaction for the 3' end cloning resulted in amplification of a fragment of

about 500 base pairs which was recovered and confirmed by DNA sequencing to be related to SEQ ID NO: 1.

The partial sequence of 549 base pairs is listed in the Sequence Listing with SEQ ID NO: 11. The corresponding deduced amino acid sequence is listed as SEQ ID NO: 12.

Claims

1. A protease of peptidase family S2A and/or peptidase family S1E and comprising an amino acid sequence which, when aligned according to Fig. 1, does not comprise any of the following:

a) (K114 + P121 + R130 + S162 + R163 + R174 + Q177), and not

b) (R114 + P121 + R130 + R162 + T163 + P174 + N177);

wherein the numbering of each amino acid residue corresponds to the numbering of the *Gliocladium* protease (amino acids 1-186 of SEQ ID NO: 6).

2. The protease of claim 1 which does not comprise any of the following:

a) (A61 + V63 + G77 + I111 + K114 + P121 + R130 + S162 + R163 + R174 + Q177); and not

b) (A61 + V63 + R77 + I111 + R114 + P121 + R130 + R162 + T163 + P174 + N177).

3. The protease of any one of claims 1-2 which comprises at least one of the following: CDEFGHIKLMNPQRSTVWY61; ACDEFGHIKLMNPQRSTWY63; ACDEFHIKLMNPQSTVWY77; ACDEFGHKLMNPQRSTVWY111; ACDEFGHILMNPQSTVWY114; ACDEFGHIKLMNQRSTVWY121; ACDEFGHIKLMNPQSTVWY130; ACDEFGHIKLMNPQSTVWY162; ACDEFGHIKLMNPQRSVWY163; ACDEFGHIKLMNQSTVWY174; and/or ACDEFGHIKLMPRSTVWY177.

4. The protease of claim 3 which comprises at least one of the following: ACDEFGHILMNPQSTVWY114; ACDEFGHIKLMNQRSTVWY121; ACDEFGHIKLMNPQSTVWY130; ACDEFGHIKLMNPQSTVWY162; ACDEFGHIKLMNPQRSVWY163; ACDEFGHIKLMNQSTVWY174; and/or ACDEFGHIKLMPRSTVWY177.

5. The protease of any one of claims 1-4 which comprises at least one of the following: S61; I63; N77 or T77; (V111 or L111); (L114 or Y114); (S121 or D121); (Q130 or M130); (N162 or T162 or S162); (S163 or R163); E174; and/or (S177 or E177).

6. The protease of claim 5 which comprises at least one of (L114 or Y114); (S121 or D121); (Q130 or M130); (N162 or T162 or S162); (S163 or R163); E174; and/or (S177 or E177).

7. The protease of claim 5, which comprises

(S61 + I63 + (N77 or T77) + (V111 or L111) + (L114 or Y114) + (S121 or D121) + (Q130 or M130) + (N162 or T162 or S162) + (S163 or R163) + E174 + (S177 or E177)).

8. The protease of claim 6, which comprises ((L114 or Y114) + (S121 or D121) + (Q130 or M130) + (N162 or T162 or S162) + (S163 or R163) + E174 + (S177 or E177)).

9. The protease of any one of the preceding claims which comprises at least one of (N162 or T162); and/or S163.

10. The protease of any one of the preceding claims which comprises (H30+D59+S140).

11. The protease of any one of the preceding claims which comprises (A28 + G29 + H30 + C31 + G32); D59; and (C134 + A135 + E136 + P137 + G138 + D139 + S140 + G141 + G142 + S143).

12. The protease of any one of the preceding claims, selected from the group consisting of:

(a) a protease having an amino acid sequence which has a degree of identity of at least 40% to amino acids 1 to 186 of SEQ ID NO: 2, amino acids 1 to 185 of SEQ ID NO: 10, amino acids 1 to 186 of SEQ ID NO: 8, amino acids 1 to 186 of SEQ ID NO: 6, and/or to amino acids 1 to 65 of SEQ ID NO: 12;

(b) a protease which is encoded by a nucleic acid sequence which hybridizes under low stringency conditions with

(i) the mature protease encoding parts of the plasmids contained in *Escherichia coli* DSM 15509, DSM 16128, DSM 16052, and/or DSM 16051;

(ii) nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1; nucleotides 652-1206, 139-1206, or 82-1206 of SEQ ID NO: 9; nucleotides 638-1195, 134-1195, or 77-1195 of SEQ ID NO: 7; nucleotides 636-1193, 114-1193, or 57-1193 of SEQ ID NO: 5; and/or nucleotides 355-549, or 1-549 of SEQ ID NO: 11;

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, or

(iv) a complementary strand of (i), (ii) or (iii);

(c) a variant of the protease having an amino acid sequence of amino acids 1 to 186, -170 to 186, or -189 to 186 of SEQ ID NO: 2; amino acids 1 to 185, 171 to 185, or 190 to 185 of SEQ ID NO: 10; amino acids 1 to 186, -168 to 186, or 187 to 186 of SEQ ID NO:

8; amino acids 1 to 186, 174 to 186, or 193 to 186 of SEQ ID NO: 6; and/or amino acids 1 to 65, or 118 to 65 of SEQ ID NO: 12; comprising a substitution, deletion, and/or insertion of one or more amino acids;

(d) an allelic variant of (a) or (b); and

(e) a fragment of (a), (b), or (d) that has protease activity.

13. An isolated nucleic acid sequence comprising a nucleic acid sequence which

(a) encodes the protease of any one of claims 1-12;

(b) encodes a protease which hybridizes under low stringency conditions with

(i) the mature protease encoding part of the plasmid contained in *Escherichia coli* DSM 15509, DSM 16128, DSM 16052, and/or DSM 16051;

(ii) nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1; nucleotides 652-1206, 139-1206, or 82-1206 of SEQ ID NO: 9; nucleotides 638-1195, 134-1195, or 77-1195 of SEQ ID NO: 7; nucleotides 636-1193, 114-1193, or 57-1193 of SEQ ID NO: 5; and/or nucleotides 355-549, or 1-549 of SEQ ID NO: 11;

(iii) a subsequence of (i) or (ii) of at least 100 nucleotides, and/or

(iv) a complementary strand of (i), (ii), or (iii); and/or

(c) encodes a protease which has a degree of identity of at least 61%, 64%, 64%, 65%, and 57%, respectively, to nucleotides 726-1283, 216-1283, or 159-1283 of SEQ ID NO: 1; nucleotides 652-1206, 139-1206, or 82-1206 of SEQ ID NO: 9; nucleotides 638-1195, 134-1195, or 77-1195 of SEQ ID NO: 7; nucleotides 636-1193, 114-1193, or 57-1193 of SEQ ID NO: 5; and/or nucleotides 355-549, or 1-549 of SEQ ID NO: 11, respectively.

14. A nucleic acid construct comprising the nucleic acid sequence of claim 13 operably linked to one or more control sequences that direct the production of the protease in a suitable expression host.

15. A recombinant expression vector comprising the nucleic acid construct of claim 14.

16. A recombinant host cell comprising the nucleic acid construct of claim 14 or the vector of claim 15.

17. A method for producing a polypeptide of any one of claims 1-12, the method comprising (a) cultivating a recombinant host cell of claim 16 to produce a supernatant comprising the protease; and (b) recovering the protease.

18. A transgenic plant, or plant part, capable of expressing the protease of any one of claims 1-12.

19. A transgenic, non-human animal, or products, or elements thereof, being capable of expressing the protease of any one of claims 1-12.

20. Use of at least one protease of any one of claims 1-12 (i) in animal feed; (ii) in the preparation of a composition for use in animal feed; (iii) for improving the nutritional value of an animal feed; (iv) for increasing digestible and/or soluble protein in animal diets; (v) for increasing the degree of hydrolysis of proteins in animal diets; and/or (vi) for the treatment of vegetable proteins.

21. An animal feed additive comprising at least one protease of any one of claims 1-12; and

(a) at least one fat-soluble vitamin, and/or

(b) at least one water-soluble vitamin, and/or

(c) at least one trace mineral.

22. An animal feed composition having a crude protein content of 50 to 800 g/kg and comprising at least one protease of any one of claims 1-12, or at least one feed additive of claim 21.

23. A detergent composition comprising at least one protease of any one of claims 1-12 and at least one surfactant.

24. Use of at least one protease having at least 55% identity to amino acids 1-186 of SEQ ID NO: 2 (i) in animal feed; (ii) in the preparation of a composition for use in animal feed; (iii) for improving the nutritional value of an animal feed; (v) for increasing digestible and/or soluble protein in animal diets; (vi) for increasing the degree of hydrolysis of proteins in animal diets; and/or (iv) for the treatment of vegetable proteins.

25. An animal feed additive comprising at least one protease as defined in claim 24; and

(a) at least one fat-soluble vitamin, and/or

(b) at least one water-soluble vitamin, and/or

(c) at least one trace mineral.

26. An animal feed composition having a crude protein content of 50 to 800 g/kg and comprising at least one protease as defined in claim 24, or at least one feed additive of claim 25.

```
                                                                SEQ ID NO:
Glíocladium      ----------VQGGDAYYINNSARCSIGFSVTTG-----FVSAGHCGSVGASATTS   41   6
Periconia CBS114000 ----------VRGGDAYLINNAARCSVGFSVSGGG----FVSAGHCGRSGNTATTT        12
Curvularia       ----------VRGGDAYYIG-GGRCSVGFSVTTG-----FVSAGHCGRAGTVATTS        10
Periconia CBS114002 ----------VRGGDAYYINTNSRCSVGFSVTTG-----FVSAGHCGSNGARATTS     8
Brachysporiella  ----------ILGGDPFVINNSAVCSVGFAVSGG-----FVSAGHCGGQGSPVTYI        2
Metarhizium      ----------ATVQGGDVYYINRSSRCSIGFAVTTG-----FVSAGHCGGSGASATTS        4
Nocardiopsis     ----------ADIIGGLAYTMG--GRCSVGFAATNAAGQPGFVTAGHCGRVGTQVTIG     26

Glíocladium      SGAALGTFSGSVFPGSADMSYIRTVSGTTLTGYINNY-SGGRLPVSGSTASAVGSSICRS  100
Periconia CBS114000 GGANIGSFAGSTFPGN-DYSYISR----------------------------------
Curvularia       GGTTLGTFSGSVFPGNGDYSYIRGTSSNTYSGTINNY-SGGSLPVSGSTASAVGSSICRS
Periconia CBS114002 SGASLGTFAGSVFPGSGDIYSFIRGSSGNTFTGTINNY-SGGTIAVSGSTAAGVGSSICRS
Brachysporiella  DGGALGTIEGSVFPGDADMSFIRAVDGTDLPGIVGTY-GNGDQPIFGSNVAPIGSGVCRS
Metarhizium      SGEALGTFSGSVFPGSADMAYVRTVSGTVLRGYINGY-GQGSFPVSGSSEAAVGASICRS
Nocardiopsis     NG--RGVFEQSVFPGN-DAAFVRGTSNFTLTNLVSRYNTGGYATVAGHNQAPIGSSVCRS

Glíocladium      GSTTGVHCGSVQALGVTVSYSQGRVTGLTQTSVCAEPGDSGGSYYTGAQAQGVTSGGSGS  160
Periconia CBS114000 ----------------------------------------------------------
Curvularia       GSTTGVHCGTVRSLGATVTYSEGRVTGLTQTNVCAEPGDSGGSFYTGAQAQGVTSGGSGD
Periconia CBS114002 GSTTGVHCGTVRSLGATVNYSQGSVTGLTQTNVCAEPGDSGGSFYSGSQAQGVTSGGSGN
Brachysporiella  GTTTGYHCGQLDAYDVTVNYDVGPVFGLTMTSACAEPGDSGGSFFAGDQAQGVTSGGSGD
Metarhizium      GSTTQVHCGTIGAKGATVNYPQGAVSGLTRTSVCAEPGDSGGSFYSGSQAQGVTSGGSGD
Nocardiopsis     GSTTGWHCGTIQARGQSVSYPEGTVTNMTRTTVCAEPGDSGGSYISGTQAQGVTSGGSGN

Glíocladium      CSSGGTTFFQPVNEILSTYGLTLVRA   186
Periconia CBS114000 ----------------------------
Curvularia       CSRGGTTYFQPVNEILSAYGLTLVRG
Periconia CBS114002 CNSGGTTFFQPVNEILSAYGLTLVRG
Brachysporiella  CTSGGQTFFQPVNEILETYGLSLTTA
Metarhizium      CSRGGTTYFQPVNRILQTYGLTLVTA
Nocardiopsis     CRTGGTTFYQEVTPMVNSWGVRLRT-
```

Fig. 1

10364204WO.ST25
SEQUENCE LISTING

<110> Novozymes A/S
Pontoppidan, Katrine
De Maria, Leonardo
Liu, Ye
Wu, Wenping
Tang, Lan

<120> Proteases

<130> 10364.204-WO

<160> 26

<170> PatentIn version 3.2

<210> 1
<211> 1400
<212> DNA
<213> Brachysporiella gayana

<220>
<221> sig_peptide
<222> (159)..(215)

<220>
<221> CDS
<222> (159)..(1283)

<220>
<221> misc_feature
<222> (216)..(725)
<223> pro_peptide

<220>
<221> mat_peptide
<222> (726)..(1283)

<400> 1
gtaggagctg tagaatcagc acatgaggca agtataaaag aaccagcatg ggatgatcaa 60

agtctgccaa ttcaaaggag caccatcaag ccgtcttgtc tagaactcct tgaacaccct 120

gtctactcca gtactcttgt cacagaacac atctagat atg gag ctc aca agc 173
Met Glu Leu Thr Ser
-185

ctc atc gcc gca ctc gca gtt att ctg cct att gcc tac ggt gtt 218
Leu Ile Ala Ala Leu Ala Val Ile Leu Pro Ile Ala Tyr Gly Val
-180 -175 -170

ccc atg gat gcc acc acc aac ctt tct ccc aag gtc ctg gcc gct 263
Pro Met Asp Ala Thr Thr Asn Leu Ser Pro Lys Val Leu Ala Ala
-165 -160 -155

atg aag cgc gac ctg gga ctt gac gcc agg gag gcc act gcc cgt 308
Met Lys Arg Asp Leu Gly Leu Asp Ala Arg Glu Ala Thr Ala Arg
-150 -145 -140

gtc acc ttc gaa cgt cgt gct ggc gat gtc atc gag gag ctg cgc 353
Val Thr Phe Glu Arg Arg Ala Gly Asp Val Ile Glu Glu Leu Arg
-135 -130 -125

agc tcc ctg gga gat tcg ttc gcc ggt gct tgg gtt acg gat ggc 398
Ser Ser Leu Gly Asp Ser Phe Ala Gly Ala Trp Val Thr Asp Gly
-120 -115 -110

```
aag gtc atc aac att  ggt gtc act gat caa  gct ttg gtc tcc aag gtt     446
Lys Val Ile Asn Ile  Gly Val Thr Asp Gln  Ala Leu Val Ser Lys Val
                -105                 -100                 -95

aag gaa gct ggc gct gaa ccg atg gtt atg aag aac agc ctc ggg aag     494
Lys Glu Ala Gly Ala Glu Pro Met Val Met Lys Asn Ser Leu Gly Lys
            -90                  -85                  -80

ctt caa gag gca aag aag aag ctt gat cag atc atc aag gag aag ccg     542
Leu Gln Glu Ala Lys Lys Lys Leu Asp Gln Ile Ile Lys Glu Lys Pro
        -75                  -70                  -65

aag acc ctc agc acc tca ggc aag ccc ggc att gca aca tac tac gtt     590
Lys Thr Leu Ser Thr Ser Gly Lys Pro Gly Ile Ala Thr Tyr Tyr Val
    -60                  -55                  -50

gac att gag acc aac aag ctc atc atc acg gca ctc tcc acc agt atc     638
Asp Ile Glu Thr Asn Lys Leu Ile Ile Thr Ala Leu Ser Thr Ser Ile
-45                  -40                  -35                  -30

act caa gct gaa gat ctg gct aag gag gtt ggc ctt tct gag tct gag     686
Thr Gln Ala Glu Asp Leu Ala Lys Glu Val Gly Leu Ser Glu Ser Glu
                -25                  -20                  -15

ttc gag gtg cgc aag act gag aag atg cca tcc cct ttc atc ctc ggc     734
Phe Glu Val Arg Lys Thr Glu Lys Met Pro Ser Pro Phe Ile Leu Gly
            -10                  -5                   -1  1

gga gac ccc ttt gtc atc aac aac agt gcc gtg tgc tct gtc ggc ttc     782
Gly Asp Pro Phe Val Ile Asn Asn Ser Ala Val Cys Ser Val Gly Phe
    5                   10                  15

gcc gtc tct ggc ggg ttt gtc tca gct ggc cac tgt ggc ggt caa ggc     830
Ala Val Ser Gly Gly Phe Val Ser Ala Gly His Cys Gly Gly Gln Gly
20                  25                  30                  35

agc cct gtc acc tat atc gac ggt ggc gca ctt gga acg atc gaa gga     878
Ser Pro Val Thr Tyr Ile Asp Gly Gly Ala Leu Gly Thr Ile Glu Gly
                40                  45                  50

tct gtc ttc ccc ggt gat gca gat atg tcc ttc atc cgt gcc gtt gac     926
Ser Val Phe Pro Gly Asp Ala Asp Met Ser Phe Ile Arg Ala Val Asp
            55                  60                  65

ggc act gac ctc cct ggc atc gtt ggt acc tat ggc aac ggt gat cag     974
Gly Thr Asp Leu Pro Gly Ile Val Gly Thr Tyr Gly Asn Gly Asp Gln
        70                  75                  80

ccc atc ttt ggc agc aat gtc gca ccc atc ggc tct ggt gtc tgc cgc    1022
Pro Ile Phe Gly Ser Asn Val Ala Pro Ile Gly Ser Gly Val Cys Arg
    85                  90                  95

tca gga aca act acc ggc tat cac tgc ggc cag ctt gat gcc tac gac    1070
Ser Gly Thr Thr Thr Gly Tyr His Cys Gly Gln Leu Asp Ala Tyr Asp
100                 105                 110                 115

gtc act gtc aac tac gac gtg gga cct gtg ttc ggt ctt acc atg acc    1118
Val Thr Val Asn Tyr Asp Val Gly Pro Val Phe Gly Leu Thr Met Thr
                120                 125                 130

tct gct tgc gct gag cct gga gac tct ggc ggc tcc ttc ttt gcc ggt    1166
Ser Ala Cys Ala Glu Pro Gly Asp Ser Gly Gly Ser Phe Phe Ala Gly
            135                 140                 145

gac cag gct cag ggc gtc acc tcg gga ggt tct ggt gat tgc acc agc    1214
Asp Gln Ala Gln Gly Val Thr Ser Gly Gly Ser Gly Asp Cys Thr Ser
        150                 155                 160
```

```
ggt ggt cag acc ttc ttc cag ccc gtg aac gag att ctg gag acc tat      1262
Gly Gly Gln Thr Phe Phe Gln Pro Val Asn Glu Ile Leu Glu Thr Tyr
    165                 170                 175

ggt ctc tcg ctc acc acg gcc taattggatg aggctttgga cagccaagag       1313
Gly Leu Ser Leu Thr Thr Ala
180                 185

cctcaacttt tcatctgtat atagcaatta ttttcgtatc tcaagtactt agatcgtcac   1373

gatcaataca tatggacttc gcttggc                                      1400
```

<210> 2
<211> 375
<212> PRT
<213> Brachysporiella gayana

<400> 2

```
Met Glu Leu Thr Ser  Leu Ile Ala Ala Leu  Ala Val Ile Leu Pro
                -185                -180                -175

Ile Ala Tyr Gly Val  Pro Met Asp Ala Thr  Thr Asn Leu Ser Pro
                -170                -165                -160

Lys Val Leu Ala Ala  Met Lys Arg Asp Leu  Gly Leu Asp Ala Arg
                -155                -150                -145

Glu Ala Thr Ala Arg  Val Thr Phe Glu Arg  Arg Ala Gly Asp Val
                -140                -135                -130

Ile Glu Glu Leu Arg  Ser Ser Leu Gly Asp  Ser Phe Ala Gly Ala
                -125                -120                -115

Trp Val Thr Asp Gly  Lys Val Ile Asn Ile  Gly Val Thr Asp Gln
                -110                -105                -100

Ala Leu Val Ser Lys Val Lys Glu Ala Gly Ala Glu Pro Met Val Met
                -95                 -90                 -85

Lys Asn Ser Leu Gly Lys Leu Gln Glu Ala Lys Lys Lys Leu Asp Gln
            -80                 -75                 -70

Ile Ile Lys Glu Lys Pro Lys Thr Leu Ser Thr Ser Gly Lys Pro Gly
        -65                 -60                 -55

Ile Ala Thr Tyr Tyr Val Asp Ile Glu Thr Asn Lys Leu Ile Ile Thr
    -50                 -45                 -40

Ala Leu Ser Thr Ser Ile Thr Gln Ala Glu Asp Leu Ala Lys Glu Val
-35                 -30                 -25                 -20

Gly Leu Ser Glu Ser Glu Phe Glu Val Arg Lys Thr Glu Lys Met Pro
                -15                 -10                 -5
```

```
Ser Pro Phe Ile Leu Gly Gly Asp Pro Phe Val Ile Asn Asn Ser Ala
       -1  1               5                  10

Val Cys Ser Val Gly Phe Ala Val Ser Gly Gly Phe Val Ser Ala Gly
    15                  20                  25

His Cys Gly Gly Gln Gly Ser Pro Val Thr Tyr Ile Asp Gly Gly Ala
30                  35                  40                  45

Leu Gly Thr Ile Glu Gly Ser Val Phe Pro Gly Asp Ala Asp Met Ser
                50                  55                  60

Phe Ile Arg Ala Val Asp Gly Thr Asp Leu Pro Gly Ile Val Gly Thr
            65                  70                  75

Tyr Gly Asn Gly Asp Gln Pro Ile Phe Gly Ser Asn Val Ala Pro Ile
        80                  85                  90

Gly Ser Gly Val Cys Arg Ser Gly Thr Thr Thr Gly Tyr His Cys Gly
    95                  100                 105

Gln Leu Asp Ala Tyr Asp Val Thr Val Asn Tyr Asp Val Gly Pro Val
110                 115                 120                 125

Phe Gly Leu Thr Met Thr Ser Ala Cys Ala Glu Pro Gly Asp Ser Gly
                130                 135                 140

Gly Ser Phe Phe Ala Gly Asp Gln Ala Gln Gly Val Thr Ser Gly Gly
            145                 150                 155

Ser Gly Asp Cys Thr Ser Gly Gly Gln Thr Phe Phe Gln Pro Val Asn
        160                 165                 170

Glu Ile Leu Glu Thr Tyr Gly Leu Ser Leu Thr Thr Ala
    175                 180                 185
```

<210> 3
<211> 1125
<212> DNA
<213> Metarhizium anisopliae

<220>
<221> CDS
<222> (1)..(1122)

<220>
<221> sig_peptide
<222> (1)..(54)

<220>
<221> misc_feature
<222> (55)..(558)
<223> pro_peptide

<220>
<221> mat_peptide
<222> (559)..(1122)

<400> 3

```
atg gag  ctt acc aaa ttt ctt  gcc tta ctg gca gtt  atc ctg ccc        45
Met Glu  Leu Thr Lys Phe Leu  Ala Leu Leu Ala Val  Ile Leu Pro
    -185                 -180                 -175

gtc gcc  tac ggt gca cca acg  cag gcg gca agc ctg  cac ccc cag        90
Val Ala  Tyr Gly Ala Pro Thr  Gln Ala Ala Ser Leu  His Pro Gln
    -170                 -165                 -160

att ttg  gag gcc atg aag cgc  gac ttg ggg ctg aac  gcc gag cag       135
Ile Leu  Glu Ala Met Lys Arg  Asp Leu Gly Leu Asn  Ala Glu Gln
    -155                 -150                 -145

gcc act  gtt cgt gtg gcg cgg  gag atc cat gcc acc  gat gtt att       180
Ala Thr  Val Arg Val Ala Arg  Glu Ile His Ala Thr  Asp Val Ile
    -140                 -135                 -130

gag cag  ctg cgc agc tca gta  gcg ttc gct ggt gct  tgg att gac       225
Glu Gln  Leu Arg Ser Ser Val  Ala Phe Ala Gly Ala  Trp Ile Asp
    -125                 -120                 -115

gcg gac  gtg cta tac atc ggc  att act gac caa gcc  ttg gcc gat       270
Ala Asp  Val Leu Tyr Ile Gly  Ile Thr Asp Gln Ala  Leu Ala Asp
    -110                 -105                 -100

gag gtc act gct gcc ggc gcc acg ccg att gtc atg acc aac agc ctg      318
Glu Val Thr Ala Ala Gly Ala Thr Pro Ile Val Met Thr Asn Ser Leu
    -95                  -90                  -85

tcc aag ctg gaa aag gcc aag gag gat ctc gat aag ata ttc atc ggc      366
Ser Lys Leu Glu Lys Ala Lys Glu Asp Leu Asp Lys Ile Phe Ile Gly
-80                  -75                  -70                  -65

cga gcc aac acc ctg gaa aca tct tcg gac act agc tct ggc att gca      414
Arg Ala Asn Thr Leu Glu Thr Ser Ser Asp Thr Ser Ser Gly Ile Ala
                -60                  -55                  -50

tcg tat ttc gtt gat gtc gcc gcc aac aag ctc gtt ata gag gct ctc      462
Ser Tyr Phe Val Asp Val Ala Ala Asn Lys Leu Val Ile Glu Ala Leu
            -45                  -40                  -35

gcc gac agt cac ggc cat gct gag caa cta gcc gcg cag gtt ggg ctt      510
Ala Asp Ser His Gly His Ala Glu Gln Leu Ala Ala Gln Val Gly Leu
        -30                  -25                  -20

aca tcc gaa ttc gag gtg cgg act gtt gag acg atg ccg act acc atg      558
Thr Ser Glu Phe Glu Val Arg Thr Val Glu Thr Met Pro Thr Thr Met
    -15                  -10                  -5                   -1

gcc acg gtt cag ggt ggt gat gtc tat tat att aat aga agc tcc cgc      606
Ala Thr Val Gln Gly Gly Asp Val Tyr Tyr Ile Asn Arg Ser Ser Arg
1                5                   10                  15

tgc tct atc ggt ttc gca gta acc aca ggt ttc gtg tcc gct gga cac      654
Cys Ser Ile Gly Phe Ala Val Thr Thr Gly Phe Val Ser Ala Gly His
            20                  25                  30

tgt gga gga tca gga gct tca gct aca aca agt agt ggt gag gcc cta      702
Cys Gly Gly Ser Gly Ala Ser Ala Thr Thr Ser Ser Gly Glu Ala Leu
        35                  40                  45

gga acc ttt tcg ggc tcc gtc ttc cct ggc agt gcc gac atg gcc tac      750
Gly Thr Phe Ser Gly Ser Val Phe Pro Gly Ser Ala Asp Met Ala Tyr
    50                  55                  60
```

```
gtc cgc act gta agt gga aca gtc ctt aga ggc tac atc aac ggc tac      798
Val Arg Thr Val Ser Gly Thr Val Leu Arg Gly Tyr Ile Asn Gly Tyr
65                  70                  75                  80

ggc caa ggg agc ttt ccc gtc tca gga agc tcc gag gct gcc gtt gga      846
Gly Gln Gly Ser Phe Pro Val Ser Gly Ser Ser Glu Ala Ala Val Gly
                85                  90                  95

gct agc atc tgc cgc tcc ggc tca acc act caa gtc cac tgc ggc acg      894
Ala Ser Ile Cys Arg Ser Gly Ser Thr Thr Gln Val His Cys Gly Thr
            100                 105                 110

att ggt gcc aag ggc gcc acg gtt aac tac cct caa gga gct gtt tcg      942
Ile Gly Ala Lys Gly Ala Thr Val Asn Tyr Pro Gln Gly Ala Val Ser
        115                 120                 125

ggc ctc act cgg act agc gtc tgc gcc gag ccc ggc gac tca ggc ggt      990
Gly Leu Thr Arg Thr Ser Val Cys Ala Glu Pro Gly Asp Ser Gly Gly
    130                 135                 140

tct ttc tac tcc ggc tcc cag gcg cag ggt gtc acc tcg gga ggc agc     1038
Ser Phe Tyr Ser Gly Ser Gln Ala Gln Gly Val Thr Ser Gly Gly Ser
145                 150                 155                 160

ggc gac tgc agc cgt gga ggc acg acc tat ttc cag cct gtt aat agg     1086
Gly Asp Cys Ser Arg Gly Gly Thr Thr Tyr Phe Gln Pro Val Asn Arg
                165                 170                 175

atc ctc cag aca tat ggc ctt acc ttg gtc acg gcg tag                 1125
Ile Leu Gln Thr Tyr Gly Leu Thr Leu Val Thr Ala
            180                 185
```

<210> 4
<211> 374
<212> PRT
<213> Metarhizium anisopliae

<400> 4

```
Met Glu  Leu Thr Lys Phe Leu  Ala Leu Leu Ala Val  Ile Leu Pro
    -185                -180                -175

Val Ala  Tyr Gly Ala Pro Thr  Gln Ala Ala Ser Leu  His Pro Gln
    -170                -165                -160

Ile Leu  Glu Ala Met Lys Arg  Asp Leu Gly Leu Asn  Ala Glu Gln
    -155                -150                -145

Ala Thr  Val Arg Val Ala Arg  Glu Ile His Ala Thr  Asp Val Ile
    -140                -135                -130

Glu Gln  Leu Arg Ser Ser Val  Ala Phe Ala Gly Ala  Trp Ile Asp
    -125                -120                -115

Ala Asp  Val Leu Tyr Ile Gly  Ile Thr Asp Gln Ala  Leu Ala Asp
    -110                -105                -100

Glu Val Thr Ala Ala Gly Ala Thr Pro Ile Val Met Thr Asn Ser Leu
    -95                 -90                 -85
```

```
Ser Lys Leu Glu Lys Ala Lys Glu Asp Leu Asp Lys Ile Phe Ile Gly
-80                 -75                 -70                 -65

Arg Ala Asn Thr Leu Glu Thr Ser Ser Asp Thr Ser Ser Gly Ile Ala
                -60                 -55                 -50

Ser Tyr Phe Val Asp Val Ala Ala Asn Lys Leu Val Ile Glu Ala Leu
            -45                 -40                 -35

Ala Asp Ser His Gly His Ala Glu Gln Leu Ala Ala Gln Val Gly Leu
        -30                 -25                 -20

Thr Ser Glu Phe Glu Val Arg Thr Val Glu Thr Met Pro Thr Thr Met
    -15                 -10                 -5                  -1

Ala Thr Val Gln Gly Gly Asp Val Tyr Tyr Ile Asn Arg Ser Ser Arg
1               5                   10                  15

Cys Ser Ile Gly Phe Ala Val Thr Thr Gly Phe Val Ser Ala Gly His
            20                  25                  30

Cys Gly Gly Ser Gly Ala Ser Ala Thr Thr Ser Ser Gly Glu Ala Leu
        35                  40                  45

Gly Thr Phe Ser Gly Ser Val Phe Pro Gly Ser Ala Asp Met Ala Tyr
    50                  55                  60

Val Arg Thr Val Ser Gly Thr Val Leu Arg Gly Tyr Ile Asn Gly Tyr
65                  70                  75                  80

Gly Gln Gly Ser Phe Pro Val Ser Gly Ser Ser Glu Ala Ala Val Gly
                85                  90                  95

Ala Ser Ile Cys Arg Ser Gly Ser Thr Thr Gln Val His Cys Gly Thr
            100                 105                 110

Ile Gly Ala Lys Gly Ala Thr Val Asn Tyr Pro Gln Gly Ala Val Ser
        115                 120                 125

Gly Leu Thr Arg Thr Ser Val Cys Ala Glu Pro Gly Asp Ser Gly Gly
    130                 135                 140

Ser Phe Tyr Ser Gly Ser Gln Ala Gln Gly Val Thr Ser Gly Gly Ser
145                 150                 155                 160

Gly Asp Cys Ser Arg Gly Gly Thr Thr Tyr Phe Gln Pro Val Asn Arg
                165                 170                 175

Ile Leu Gln Thr Tyr Gly Leu Thr Leu Val Thr Ala
            180                 185
```

<210> 5
<211> 1279
<212> DNA
<213> Gliocladium sp. CBS 114001

<220>
<221> sig_peptide
<222> (57)..(113)

<220>
<221> CDS
<222> (57)..(1193)

<220>
<221> mat_peptide
<222> (636)..(1193)

<400> 5
agcatctaca acatcgacac ttttatctat tacatcttgt ctttggaacg ttagaa atg 59
Met

gag ctc act agg ttc ctc gct gcc ctg gca ttc atc ctg ccc ctt 104
Glu Leu Thr Arg Phe Leu Ala Ala Leu Ala Phe Ile Leu Pro Leu
-190 -185 -180

gct tac agt gct ccc act gag act gtt aca agg ctg cac cca gag 149
Ala Tyr Ser Ala Pro Thr Glu Thr Val Thr Arg Leu His Pro Glu
-175 -170 -165

atc cta gct gcc atg aag cgg gac ttg ggc ctg gac act gag cag 194
Ile Leu Ala Ala Met Lys Arg Asp Leu Gly Leu Asp Thr Glu Gln
-160 -155 -150

gct act gct cgt gtt gcc cgg gag atc cag gct gcc gat gtc att 239
Ala Thr Ala Arg Val Ala Arg Glu Ile Gln Ala Ala Asp Val Ile
-145 -140 -135

gac cag ctg cgc aac tcc gtg ggt gct caa ttc gcc ggt gct tgg 284
Asp Gln Leu Arg Asn Ser Val Gly Ala Gln Phe Ala Gly Ala Trp
-130 -125 -120

gtt tcc aac gat gga aac acc atc aac gtc ggt atc act gac gag 329
Val Ser Asn Asp Gly Asn Thr Ile Asn Val Gly Ile Thr Asp Glu
-115 -110 -105

tcc ttg act gct gaa gtg act gct gct ggt gcc act cct gtc gtc ttg 377
Ser Leu Thr Ala Glu Val Thr Ala Ala Gly Ala Thr Pro Val Val Leu
-100 -95 -90

gcc aac agc ttc acc aag ctt gaa gat gcc aag gtg gca ctt gac aac 425
Ala Asn Ser Phe Thr Lys Leu Glu Asp Ala Lys Val Ala Leu Asp Asn
-85 -80 -75

ctg gac att gac cag gcg agt gct ctc gac acg acc tca tcc ggg tcg 473
Leu Asp Ile Asp Gln Ala Ser Ala Leu Asp Thr Thr Ser Ser Gly Ser
-70 -65 -60 -55

ggc att gcc tcg tac tac gtt gat gtt gtg gcc aac aag ctc atc ctg 521
Gly Ile Ala Ser Tyr Tyr Val Asp Val Val Ala Asn Lys Leu Ile Leu
-50 -45 -40

gag gca ctg gca gac agc acc gcc cat gcc gag cag ctc gcc gca cag 569
Glu Ala Leu Ala Asp Ser Thr Ala His Ala Glu Gln Leu Ala Ala Gln
-35 -30 -25

```
gtt ggg ctc act ccc tct gag ttt gaa gtg cgc gtg gtt gag tca ctc      617
Val Gly Leu Thr Pro Ser Glu Phe Glu Val Arg Val Val Glu Ser Leu
        -20                 -15                 -10

cct gag act ttc gcc aca gtc cag ggc ggc gac gcc tac tac atc aac      665
Pro Glu Thr Phe Ala Thr Val Gln Gly Gly Asp Ala Tyr Tyr Ile Asn
    -5              -1  1                5                  10

aac agc gct cgt tgc tct att ggc ttt tcc gtc act acc ggc ttt gtt      713
Asn Ser Ala Arg Cys Ser Ile Gly Phe Ser Val Thr Thr Gly Phe Val
                15                  20                  25

tcg gct ggc cac tgc ggc agc gta ggt gct tct gcc act act tcc agt      761
Ser Ala Gly His Cys Gly Ser Val Gly Ala Ser Ala Thr Thr Ser Ser
            30                  35                  40

ggt gcg gcc ctt gga acc ttc tct ggt tcc gtc ttc cca ggc agc gcc      809
Gly Ala Ala Leu Gly Thr Phe Ser Gly Ser Val Phe Pro Gly Ser Ala
        45                  50                  55

gac atg agc tac atc cgc act gtc agt gga acc acc ctc acg ggt tac      857
Asp Met Ser Tyr Ile Arg Thr Val Ser Gly Thr Thr Leu Thr Gly Tyr
    60                  65                  70

atc aac aac tac agc ggc ggt aga ctc cct gtc tct ggc agc acc gcg      905
Ile Asn Asn Tyr Ser Gly Gly Arg Leu Pro Val Ser Gly Ser Thr Ala
75                  80                  85                  90

tct gcg gtt gga tcc agc atc tgc cgc tcc ggc tct act act ggt gtc      953
Ser Ala Val Gly Ser Ser Ile Cys Arg Ser Gly Ser Thr Thr Gly Val
                95                  100                 105

cac tgc ggc agc gtc cag gcc ctc ggt gtt act gtc agc tac tcc cag     1001
His Cys Gly Ser Val Gln Ala Leu Gly Val Thr Val Ser Tyr Ser Gln
            110                 115                 120

ggc cgt gtc acc ggc ctc acc cag acc agt gtc tgc gcc gag ccc ggc     1049
Gly Arg Val Thr Gly Leu Thr Gln Thr Ser Val Cys Ala Glu Pro Gly
        125                 130                 135

gac tcc ggc ggt tcc tac tac acc ggc gcc cag gct cag ggt gtc acc     1097
Asp Ser Gly Gly Ser Tyr Tyr Thr Gly Ala Gln Ala Gln Gly Val Thr
    140                 145                 150

tca gga ggt tcc ggc tcc tgc tcc tca ggt ggt aca acc ttt ttc cag     1145
Ser Gly Gly Ser Gly Ser Cys Ser Ser Gly Gly Thr Thr Phe Phe Gln
155                 160                 165                 170

cca gtc aac gag atc ctg tcc acc tac ggt ctc aca ctg gtc cga gcc     1193
Pro Val Asn Glu Ile Leu Ser Thr Tyr Gly Leu Thr Leu Val Arg Ala
                175                 180                 185

taaataaaca gccaattaac tcaatggttg acgtacctaa gtaaagcttg aaagtattgt   1253

gactttacat atgggcaatg gttagt                                       1279
```

<210> 6
<211> 379
<212> PRT
<213> Gliocladium sp. CBS 114001

<400> 6

```
Met Glu Leu Thr  Arg Phe Leu Ala Ala  Leu Ala Phe Ile Leu  Pro
           -190                -185                -180
```

```
Leu Ala Tyr Ser  Ala Pro Thr Glu Thr  Val Thr Arg Leu His  Pro
           -175                -170                -165

Glu Ile Leu Ala  Ala Met Lys Arg Asp  Leu Gly Leu Asp Thr  Glu
           -160                -155                -150

Gln Ala Thr Ala  Arg Val Ala Arg Glu  Ile Gln Ala Ala Asp  Val
           -145                -140                -135

Ile Asp Gln Leu  Arg Asn Ser Val Gly  Ala Gln Phe Ala Gly  Ala
           -130                -125                -120

Trp Val Ser Asn  Asp Gly Asn Thr Ile  Asn Val Gly Ile Thr  Asp
           -115                -110                -105

Glu Ser Leu Thr  Ala Glu Val Thr Ala Ala Gly Ala Thr Pro Val Val
           -100                -95                 -90

Leu Ala Asn Ser Phe Thr Lys Leu Glu Asp Ala Lys Val Ala Leu Asp
        -85                 -80                 -75

Asn Leu Asp Ile Asp Gln Ala Ser Ala Leu Asp Thr Thr Ser Ser Gly
    -70                 -65                 -60

Ser Gly Ile Ala Ser Tyr Tyr Val Asp Val Val Ala Asn Lys Leu Ile
-55                 -50                 -45                 -40

Leu Glu Ala Leu Ala Asp Ser Thr Ala His Ala Glu Gln Leu Ala Ala
               -35                 -30                 -25

Gln Val Gly Leu Thr Pro Ser Glu Phe Glu Val Arg Val Val Glu Ser
           -20                 -15                 -10

Leu Pro Glu Thr Phe Ala Thr Val Gln Gly Gly Asp Ala Tyr Tyr Ile
        -5              -1  1               5

Asn Asn Ser Ala Arg Cys Ser Ile Gly Phe Ser Val Thr Thr Gly Phe
10                  15                  20                  25

Val Ser Ala Gly His Cys Gly Ser Val Gly Ala Ser Ala Thr Thr Ser
               30                  35                  40

Ser Gly Ala Ala Leu Gly Thr Phe Ser Gly Ser Val Phe Pro Gly Ser
           45                  50                  55

Ala Asp Met Ser Tyr Ile Arg Thr Val Ser Gly Thr Thr Leu Thr Gly
        60                  65                  70

Tyr Ile Asn Asn Tyr Ser Gly Gly Arg Leu Pro Val Ser Gly Ser Thr
    75                  80                  85
```

Ala Ser Ala Val Gly Ser Ser Ile Cys Arg Ser Gly Ser Thr Thr Gly
90 95 100 105

Val His Cys Gly Ser Val Gln Ala Leu Gly Val Thr Val Ser Tyr Ser
110 115 120

Gln Gly Arg Val Thr Gly Leu Thr Gln Thr Ser Val Cys Ala Glu Pro
125 130 135

Gly Asp Ser Gly Gly Ser Tyr Tyr Thr Gly Ala Gln Ala Gln Gly Val
140 145 150

Thr Ser Gly Gly Ser Gly Ser Cys Ser Ser Gly Gly Thr Thr Phe Phe
155 160 165

Gln Pro Val Asn Glu Ile Leu Ser Thr Tyr Gly Leu Thr Leu Val Arg
170 175 180 185

Ala

<210> 7
<211> 1297
<212> DNA
<213> Periconia sp. CBS 114002

<220>
<221> sig_peptide
<222> (77)..(133)

<220>
<221> CDS
<222> (77)..(1195)

<220>
<221> mat_peptide
<222> (638)..(1195)

<400> 7
agtctcatct tgctacgatc tcgtctctgc aacatcaaag gcctttctct cgcttgtctt 60

cttcacgtaa atcatc atg gag ctc acc aaa ttc ctc gct gct atg gca 109
Met Glu Leu Thr Lys Phe Leu Ala Ala Met Ala
-185 -180

gtc ttc ctg ccc atc gtt tat ggt gcc cct act cag gct gcc aac 154
Val Phe Leu Pro Ile Val Tyr Gly Ala Pro Thr Gln Ala Ala Asn
-175 -170 -165

agc ctc cac ccc gag att ttg gct gcc atg aag cgt gac ttg gga 199
Ser Leu His Pro Glu Ile Leu Ala Ala Met Lys Arg Asp Leu Gly
-160 -155 -150

ctt gac gcc gag cag gcc cat gtt cgt gtt gcc cga gag ctc aag 244
Leu Asp Ala Glu Gln Ala His Val Arg Val Ala Arg Glu Leu Lys
-145 -140 -135

gct acc gag gtg atc gag cag ttg aag acc aag gct ggc tac agc 289
Ala Thr Glu Val Ile Glu Gln Leu Lys Thr Lys Ala Gly Tyr Ser
-130 -125 -120

```
ggc gca  tgg ctc gtc gat ggt  gag ctc aac gtc gct  gtc act gac       334
Gly Ala  Trp Leu Val Asp Gly  Glu Leu Asn Val Ala  Val Thr Asp
    -115                 -110                 -105

gag gct  ttg tct tcg gag gtg acc gct gct ggc gcc aaa ccc gtc acc     382
Glu Ala  Leu Ser Ser Glu Val Thr Ala Ala Gly Ala Lys Pro Val Thr
    -100                 -95                  -90

gtc tcc aac gag ctc tct aag ctc gag gag gcc aag gag gct ctt gac     430
Val Ser Asn Glu Leu Ser Lys Leu Glu Glu Ala Lys Glu Ala Leu Asp
-85                  -80                  -75                  -70

aag ctc gat gtc ccc aaa act cta agc aag agc gct gag gag act ggc     478
Lys Leu Asp Val Pro Lys Thr Leu Ser Lys Ser Ala Glu Glu Thr Gly
                -65                  -60                  -55

atc gcc tcg tgg tac gtc gat gtc acc gcc aac aag ctc gtt ctt gag     526
Ile Ala Ser Trp Tyr Val Asp Val Thr Ala Asn Lys Leu Val Leu Glu
            -50                  -45                  -40

gca ctt tct gcc agc aca tct cag gct tcc gag cta gcc aag caa gtc     574
Ala Leu Ser Ala Ser Thr Ser Gln Ala Ser Glu Leu Ala Lys Gln Val
        -35                  -30                  -25

ggt ctt gtg gac tct gaa ttc gag gtt cgc acc gtc aac gag atg ccc     622
Gly Leu Val Asp Ser Glu Phe Glu Val Arg Thr Val Asn Glu Met Pro
    -20                  -15                  -10

acc acc ttc gcc aca gtt cgc ggt ggt gac gcc tac tac atc aac acc     670
Thr Thr Phe Ala Thr Val Arg Gly Gly Asp Ala Tyr Tyr Ile Asn Thr
-5              -1  1                5                   10

aac tcc cga tgc tct gtt ggt ttc tca gtc act acc ggt ttc gtc tct     718
Asn Ser Arg Cys Ser Val Gly Phe Ser Val Thr Thr Gly Phe Val Ser
            15                   20                  25

gct ggt cac tgc ggc agc aac ggt gcc agg gcc acc act agc agc ggt     766
Ala Gly His Cys Gly Ser Asn Gly Ala Arg Ala Thr Thr Ser Ser Gly
        30                   35                  40

gct tct ttg gga act ttc gct gga tca gta ttc ccc gga agc ggt gac     814
Ala Ser Leu Gly Thr Phe Ala Gly Ser Val Phe Pro Gly Ser Gly Asp
    45                   50                  55

tac agc ttc atc cgt ggc agc agc gga aac act ttc act ggc acc atc     862
Tyr Ser Phe Ile Arg Gly Ser Ser Gly Asn Thr Phe Thr Gly Thr Ile
60                  65                  70                  75

aac aac tac agc ggt ggt acc atc gcc gtc tca ggc agc acc gct gcc     910
Asn Asn Tyr Ser Gly Gly Thr Ile Ala Val Ser Gly Ser Thr Ala Ala
                80                  85                  90

ggc gtc ggt tcc agc atc tgc cgt tcc ggc tct acc act ggt gtt cac     958
Gly Val Gly Ser Ser Ile Cys Arg Ser Gly Ser Thr Thr Gly Val His
            95                  100                 105

tgc ggt acc gtc cgc tcc ctg ggc gca acc gtc aac tac tct cag ggc     1006
Cys Gly Thr Val Arg Ser Leu Gly Ala Thr Val Asn Tyr Ser Gln Gly
        110                 115                 120

agt gtc acc ggt ctg act cag acc aat gtc tgc gct gag ccc gga gac     1054
Ser Val Thr Gly Leu Thr Gln Thr Asn Val Cys Ala Glu Pro Gly Asp
    125                 130                 135

tct ggc ggt tcc ttc tac tct ggt tcc caa gcc cag ggt gtc acc tcc     1102
Ser Gly Gly Ser Phe Tyr Ser Gly Ser Gln Ala Gln Gly Val Thr Ser
140                 145                 150                 155
```

```
ggt ggc tct ggc aac tgc aac agc ggt gga acc acc ttc ttc cag ccc     1150
Gly Gly Ser Gly Asn Cys Asn Ser Gly Gly Thr Thr Phe Phe Gln Pro
                160                 165                 170

gtt aat gag att ctt tcc gcc tat ggc ctc act ctt gtc cgt ggt         1195
Val Asn Glu Ile Leu Ser Ala Tyr Gly Leu Thr Leu Val Arg Gly
            175                 180                 185

taagtagtta gttgttgaac tgatcgatgg tgctgttacg ggaatcccgc ggccatggcg  1255

gccgggagca tgcgacgtcg ggcccaattc gccctatagt ga                     1297
```

<210> 8
<211> 373
<212> PRT
<213> Periconia sp. CBS 114002

<400> 8

```
Met Glu Leu Thr Lys Phe Leu Ala Ala Met Ala Val Phe Leu Pro
        -185                 -180                 -175

Ile Val Tyr Gly Ala Pro Thr Gln Ala Ala Asn Ser Leu His Pro
        -170                 -165                 -160

Glu Ile Leu Ala Ala Met Lys Arg Asp Leu Gly Leu Asp Ala Glu
        -155                 -150                 -145

Gln Ala His Val Arg Val Ala Arg Glu Leu Lys Ala Thr Glu Val
        -140                 -135                 -130

Ile Glu Gln Leu Lys Thr Lys Ala Gly Tyr Ser Gly Ala Trp Leu
        -125                 -120                 -115

Val Asp Gly Glu Leu Asn Val Ala Val Thr Asp Glu Ala Leu Ser
        -110                 -105                 -100

Ser Glu Val Thr Ala Ala Gly Ala Lys Pro Val Thr Val Ser Asn Glu
        -95                  -90                  -85

Leu Ser Lys Leu Glu Glu Ala Lys Glu Ala Leu Asp Lys Leu Asp Val
    -80                  -75                  -70

Pro Lys Thr Leu Ser Lys Ser Ala Glu Glu Thr Gly Ile Ala Ser Trp
-65                  -60                  -55                  -50

Tyr Val Asp Val Thr Ala Asn Lys Leu Val Leu Glu Ala Leu Ser Ala
                -45                  -40                  -35

Ser Thr Ser Gln Ala Ser Glu Leu Ala Lys Gln Val Gly Leu Val Asp
            -30                  -25                  -20

Ser Glu Phe Glu Val Arg Thr Val Asn Glu Met Pro Thr Thr Phe Ala
        -15                  -10                  -5
```

```
Thr Val Arg Gly Gly Asp Ala Tyr Tyr Ile Asn Thr Asn Ser Arg Cys
-1  1               5                  10                  15

Ser Val Gly Phe Ser Val Thr Thr Gly Phe Val Ser Ala Gly His Cys
            20                  25                  30

Gly Ser Asn Gly Ala Arg Ala Thr Thr Ser Ser Gly Ala Ser Leu Gly
        35                  40                  45

Thr Phe Ala Gly Ser Val Phe Pro Gly Ser Gly Asp Tyr Ser Phe Ile
    50                  55                  60

Arg Gly Ser Ser Gly Asn Thr Phe Thr Gly Thr Ile Asn Asn Tyr Ser
65                  70                  75

Gly Gly Thr Ile Ala Val Ser Gly Ser Thr Ala Ala Gly Val Gly Ser
80                  85                  90                  95

Ser Ile Cys Arg Ser Gly Ser Thr Thr Gly Val His Cys Gly Thr Val
            100                 105                 110

Arg Ser Leu Gly Ala Thr Val Asn Tyr Ser Gln Gly Ser Val Thr Gly
        115                 120                 125

Leu Thr Gln Thr Asn Val Cys Ala Glu Pro Gly Asp Ser Gly Gly Ser
    130                 135                 140

Phe Tyr Ser Gly Ser Gln Ala Gln Gly Val Thr Ser Gly Gly Ser Gly
145                 150                 155

Asn Cys Asn Ser Gly Gly Thr Thr Phe Phe Gln Pro Val Asn Glu Ile
160                 165                 170                 175

Leu Ser Ala Tyr Gly Leu Thr Leu Val Arg Gly
            180                 185
```

<210> 9
<211> 1259
<212> DNA
<213> Curvularia lunata

<220>
<221> sig_peptide
<222> (82)..(138)

<220>
<221> CDS
<222> (82)..(1206)

<220>
<221> mat_peptide
<222> (652)..(1206)

<400> 9

10364204WO.ST25

```
gcagcaacta caacttccta cgatccagcc actacaacat tgactttctg tcttttctgt

gatatctcca caaccctgac a atg  gag ctc acc aaa ttc  ctt act gct ctg    111
                       Met  Glu Leu Thr Lys Phe  Leu Thr Ala Leu
                       -190                -185

gca  gtc ttt ctg ccc atc  gtc tac ggt gcc cct  acc acg act ggc      156
Ala  Val Phe Leu Pro Ile  Val Tyr Gly Ala Pro  Thr Thr Thr Gly
-180                -175                -170

aac  agc ctc cac ccc gag  atc ctg gct gcc atg  aag cgc gac ttg      201
Asn  Ser Leu His Pro Glu  Ile Leu Ala Ala Met  Lys Arg Asp Leu
-165                -160                -155

gga  ctt gat gct gag cag  gcc cac gcc cgt gtt  gca cgg gag atc      246
Gly  Leu Asp Ala Glu Gln  Ala His Ala Arg Val  Ala Arg Glu Ile
-150                -145                -140

cag  gcc act gag gtc atc  gaa cag ctc cgc acc  aag gct ggc aac      291
Gln  Ala Thr Glu Val Ile  Glu Gln Leu Arg Thr  Lys Ala Gly Asn
-135                -130                -125

tcc  ttc ggt ggt gcc tgg  ctc gtc gac ggt gag  ctc aag gtc gcc      336
Ser  Phe Gly Gly Ala Trp  Leu Val Asp Gly Glu  Leu Lys Val Ala
-120                -115                -110

gtc  act gac gag gcc ttg  acc tct gag gtc act tct tcc gga gcc aca  384
Val  Thr Asp Glu Ala Leu  Thr Ser Glu Val Thr Ser Ser Gly Ala Thr
-105                -100                -95                 -90

gca ttg gtt gtt gac acc ccc ctc tcc aag ctc cag gag gcc cag aag    432
Ala Leu Val Val Asp Thr Pro Leu Ser Lys Leu Gln Glu Ala Gln Lys
               -85                 -80                 -75

gca ctt gac aag ctg gat att gct tca gct gct ggc aag cgc tct gag    480
Ala Leu Asp Lys Leu Asp Ile Ala Ser Ala Ala Gly Lys Arg Ser Glu
          -70                 -65                 -60

gct gcc aat gga att gca gca tac tac gtc gat gtt gcc gcc aac aag    528
Ala Ala Asn Gly Ile Ala Ala Tyr Tyr Val Asp Val Ala Ala Asn Lys
     -55                 -50                 -45

ctc gtc att gag gcc ctc gag agc agc acc gca cag gct cag gag ttg    576
Leu Val Ile Glu Ala Leu Glu Ser Ser Thr Ala Gln Ala Gln Glu Leu
-40                 -35                 -30

gcc aag aag gtt ggc ctc acc gag tct gag ttt gag gtc cgc acc gtt    624
Ala Lys Lys Val Gly Leu Thr Glu Ser Glu Phe Glu Val Arg Thr Val
-25                 -20                 -15                 -10

cag gac atg ccc acc cct ctt gct acc gtt cgt ggt ggc gac gca tac    672
Gln Asp Met Pro Thr Pro Leu Ala Thr Val Arg Gly Gly Asp Ala Tyr
               -5                  -1  1                   5

tac att ggc ggc gga cga tgc tct gta ggt ttc tcc gtc acc act ggt    720
Tyr Ile Gly Gly Gly Arg Cys Ser Val Gly Phe Ser Val Thr Thr Gly
          10                  15                  20

ttc gta tct gct ggt cac tgc ggc cgt gct ggc acc gtt gct acc acc    768
Phe Val Ser Ala Gly His Cys Gly Arg Ala Gly Thr Val Ala Thr Thr
     25                  30                  35

agc ggt ggt act acc ctg gga acg ttc tcc ggc tcc gtt ttc ccc gga    816
Ser Gly Gly Thr Thr Leu Gly Thr Phe Ser Gly Ser Val Phe Pro Gly
40                  45                  50                  55

aac ggt gac tac agc tac att cgt ggt acc agc tca aac acc tac tct    864
Asn Gly Asp Tyr Ser Tyr Ile Arg Gly Thr Ser Ser Asn Thr Tyr Ser
```

60 65 70

ggc acc atc aac aac tac agc ggt ggt tcc ctc ccc gtt tct ggc agc 912
Gly Thr Ile Asn Asn Tyr Ser Gly Gly Ser Leu Pro Val Ser Gly Ser
75 80 85

acc gcc tct gcc gtt gga tcc agc atc tgc cgc tct gga tct acc act 960
Thr Ala Ser Ala Val Gly Ser Ser Ile Cys Arg Ser Gly Ser Thr Thr
90 95 100

ggt gtc cac tgc ggc act gtc cgc tct ctg gga gct acc gtc acc tac 1008
Gly Val His Cys Gly Thr Val Arg Ser Leu Gly Ala Thr Val Thr Tyr
105 110 115

tct gag ggc cgt gtt act ggc ctc act cag acc aac gtc tgc gcc gag 1056
Ser Glu Gly Arg Val Thr Gly Leu Thr Gln Thr Asn Val Cys Ala Glu
120 125 130 135

ccc ggt gac tct ggc ggt tcc ttc tac act ggt gct cag gcc caa ggt 1104
Pro Gly Asp Ser Gly Gly Ser Phe Tyr Thr Gly Ala Gln Ala Gln Gly
140 145 150

gtc acc tcc ggt ggc tcc ggt gac tgc agc agg gga ggc act acc tac 1152
Val Thr Ser Gly Gly Ser Gly Asp Cys Ser Arg Gly Gly Thr Thr Tyr
155 160 165

ttc cag ccc gtc aac gag atc ctt tcg gcc tac ggc ctt act ctg gtt 1200
Phe Gln Pro Val Asn Glu Ile Leu Ser Ala Tyr Gly Leu Thr Leu Val
170 175 180

cgc ggt taaataacta ttcttttgta tatatttcca cccttgctaa ggtggatgct 1256
Arg Gly
185

gat 1259

<210> 10
<211> 375
<212> PRT
<213> Curvularia lunata

<400> 10

Met Glu Leu Thr Lys Phe Leu Thr Ala Leu Ala Val Phe Leu Pro
-190 -185 -180

Ile Val Tyr Gly Ala Pro Thr Thr Thr Gly Asn Ser Leu His Pro
-175 -170 -165

Glu Ile Leu Ala Ala Met Lys Arg Asp Leu Gly Leu Asp Ala Glu
-160 -155 -150

Gln Ala His Ala Arg Val Ala Arg Glu Ile Gln Ala Thr Glu Val
-145 -140 -135

Ile Glu Gln Leu Arg Thr Lys Ala Gly Asn Ser Phe Gly Gly Ala
-130 -125 -120

Trp Leu Val Asp Gly Glu Leu Lys Val Ala Val Thr Asp Glu Ala
-115 -110 -105

10364204WO.ST25

```
Leu  Thr Ser Glu Val Thr Ser Ser Gly Ala Thr Ala Leu Val Val Asp
-100                 -95                 -90                 -85

Thr Pro Leu Ser Lys Leu Gln Glu Ala Gln Lys Ala Leu Asp Lys Leu
                -80                 -75                 -70

Asp Ile Ala Ser Ala Ala Gly Lys Arg Ser Glu Ala Ala Asn Gly Ile
            -65                 -60                 -55

Ala Ala Tyr Tyr Val Asp Val Ala Ala Asn Lys Leu Val Ile Glu Ala
        -50                 -45                 -40

Leu Glu Ser Ser Thr Ala Gln Ala Gln Glu Leu Ala Lys Lys Val Gly
    -35                 -30                 -25

Leu Thr Glu Ser Glu Phe Glu Val Arg Thr Val Gln Asp Met Pro Thr
-20                 -15                 -10                  -5

Pro Leu Ala Thr Val Arg Gly Gly Asp Ala Tyr Tyr Ile Gly Gly Gly
            -1  1               5                  10

Arg Cys Ser Val Gly Phe Ser Val Thr Thr Gly Phe Val Ser Ala Gly
        15                  20                  25

His Cys Gly Arg Ala Gly Thr Val Ala Thr Thr Ser Gly Gly Thr Thr
    30                  35                  40

Leu Gly Thr Phe Ser Gly Ser Val Phe Pro Gly Asn Gly Asp Tyr Ser
45                  50                  55                  60

Tyr Ile Arg Gly Thr Ser Ser Asn Thr Tyr Ser Gly Thr Ile Asn Asn
                65                  70                  75

Tyr Ser Gly Gly Ser Leu Pro Val Ser Gly Ser Thr Ala Ser Ala Val
            80                  85                  90

Gly Ser Ser Ile Cys Arg Ser Gly Ser Thr Thr Gly Val His Cys Gly
        95                  100                 105

Thr Val Arg Ser Leu Gly Ala Thr Val Thr Tyr Ser Glu Gly Arg Val
    110                 115                 120

Thr Gly Leu Thr Gln Thr Asn Val Cys Ala Glu Pro Gly Asp Ser Gly
125                 130                 135                 140

Gly Ser Phe Tyr Thr Gly Ala Gln Ala Gln Gly Val Thr Ser Gly Gly
                145                 150                 155

Ser Gly Asp Cys Ser Arg Gly Gly Thr Thr Tyr Phe Gln Pro Val Asn
            160                 165                 170
```

Glu Ile Leu Ser Ala Tyr Gly Leu Thr Leu Val Arg Gly
175 180 185

<210> 11
<211> 549
<212> DNA
<213> Periconia sp. CBS 114000

<220>
<221> CDS
<222> (1)..(549)

<220>
<221> mat_peptide
<222> (355)..(549)

<400> 11

```
ttc ggt ggt gct  tgg ttc gtc gac ggc  gag ctc aac gtt gct  gtc        45
Phe Gly Gly Ala  Trp Phe Val Asp Gly  Glu Leu Asn Val Ala  Val
          -115                 -110                 -105

act gac gag gct  ttg agt tcc gag gtg acc gct gcc ggt gct acg gcc      93
Thr Asp Glu Ala  Leu Ser Ser Glu Val Thr Ala Ala Gly Ala Thr Ala
          -100                 -95                  -90

aag gtc gtt gat acc ccc ctc tcc aag ctt gag gat gcc ttg aag gcc      141
Lys Val Val Asp Thr Pro Leu Ser Lys Leu Glu Asp Ala Leu Lys Ala
       -85                  -80                  -75

ctt gac gaa ttg gac act gtt tcg acc ctt tcc gag ggc tcc gag aaa      189
Leu Asp Glu Leu Asp Thr Val Ser Thr Leu Ser Glu Gly Ser Glu Lys
    -70                  -65                  -60

tcc gga atc gct gcc tac tac gtc gac gtt acc gcc aac aag ctc gtt      237
Ser Gly Ile Ala Ala Tyr Tyr Val Asp Val Thr Ala Asn Lys Leu Val
-55                  -50                  -45                  -40

ttg gaa tcc ctt tct gca agc act act aag gct gct gag ctt gcc aag      285
Leu Glu Ser Leu Ser Ala Ser Thr Thr Lys Ala Ala Glu Leu Ala Lys
                -35                  -30                  -25

cag gtc ggt ctc gcc gag tcc gag ttc gag gtt aag acc gtt gaa gag      333
Gln Val Gly Leu Ala Glu Ser Glu Phe Glu Val Lys Thr Val Glu Glu
             -20                  -15                  -10

ctc cca acc acc ttc gct acg gtc cgt ggc ggc gat gct tat ctc atc      381
Leu Pro Thr Thr Phe Ala Thr Val Arg Gly Gly Asp Ala Tyr Leu Ile
       -5                   -1  1                   5

aac aac gcc gct cgt tgc tct gtc ggc ttc tca gtt tct ggc gga ggc      429
Asn Asn Ala Ala Arg Cys Ser Val Gly Phe Ser Val Ser Gly Gly Gly
10                  15                  20                  25

ttc gta tct gcc ggc cac tgt ggc cgt tcc gga aac act gct acc acc      477
Phe Val Ser Ala Gly His Cys Gly Arg Ser Gly Asn Thr Ala Thr Thr
                30                  35                  40

acc ggt gga gct aac atc gga tca ttc gcg ggt tcc acc ttc ccc gga      525
Thr Gly Gly Ala Asn Ile Gly Ser Phe Ala Gly Ser Thr Phe Pro Gly
            45                  50                  55

aac gac tac agc tac att tcg cgg                                     549
Asn Asp Tyr Ser Tyr Ile Ser Arg
        60                  65
```

10364204WO.ST25

<210> 12
<211> 183
<212> PRT
<213> Periconia sp. CBS 114000

<400> 12

Phe Gly Gly Ala Trp Phe Val Asp Gly Glu Leu Asn Val Ala Val
-115 -110 -105

Thr Asp Glu Ala Leu Ser Ser Glu Val Thr Ala Ala Gly Ala Thr Ala
-100 -95 -90

Lys Val Val Asp Thr Pro Leu Ser Lys Leu Glu Asp Ala Leu Lys Ala
-85 -80 -75

Leu Asp Glu Leu Asp Thr Val Ser Thr Leu Ser Glu Gly Ser Glu Lys
-70 -65 -60

Ser Gly Ile Ala Ala Tyr Tyr Val Asp Val Thr Ala Asn Lys Leu Val
-55 -50 -45 -40

Leu Glu Ser Leu Ser Ala Ser Thr Thr Lys Ala Ala Glu Leu Ala Lys
-35 -30 -25

Gln Val Gly Leu Ala Glu Ser Glu Phe Glu Val Lys Thr Val Glu Glu
-20 -15 -10

Leu Pro Thr Thr Phe Ala Thr Val Arg Gly Gly Asp Ala Tyr Leu Ile
-5 -1 1 5

Asn Asn Ala Ala Arg Cys Ser Val Gly Phe Ser Val Ser Gly Gly Gly
10 15 20 25

Phe Val Ser Ala Gly His Cys Gly Arg Ser Gly Asn Thr Ala Thr Thr
30 35 40

Thr Gly Gly Ala Asn Ile Gly Ser Phe Ala Gly Ser Thr Phe Pro Gly
45 50 55

Asn Asp Tyr Ser Tyr Ile Ser Arg
60 65

<210> 13
<211> 17
<212> DNA
<213> Artificial

<220>
<223> Primer

<220>
<221> misc_feature
<222> (3)..(3)
<223> N in position 3 means A, C, T and G

<220>
<221> misc_feature
<222> (9)..(9)
<223> N in position 9 means A and G

<220>
<221> misc_feature
<222> (12)..(12)
<223> N in position 12 means C, G, A and T

<220>
<221> misc_feature
<222> (15)..(16)
<223> N in position 15 and 16 means T and C

<400> 13
gcnatgaanc gngannt 17

<210> 14
<211> 19
<212> DNA
<213> Artificial

<220>
<223> Primer

<220>
<221> misc_feature
<222> (3)..(3)
<223> N in position 3 means G, T and C

<220>
<221> misc_feature
<222> (5)..(5)
<223> N in position 5 means A and T

<220>
<221> misc_feature
<222> (6)..(6)
<223> N in position 6 means G and T

<220>
<221> misc_feature
<222> (9)..(9)
<223> N in position 9 means G and T

<220>
<221> misc_feature
<222> (12)..(12)
<223> N in position 12 means C and G

<220>
<221> misc_feature
<222> (18)..(18)
<223> N in position 18 means C and G

<400> 14
gtngnnccng ancggcana 19

<210> 15
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 15
actgtggcga aagtctca 18

<210> 16
<211> 17
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 16
cgcacttcaa actcaga 17

<210> 17
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 17
gttgtcaagt gccacctt 18

<210> 18
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 18
tcgttgctct attggctt 18

<210> 19
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 19
cttctctggt tccgtctt 18

<210> 20
<211> 17
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 20
actaaccatt gcccata 17

<210> 21

<211> 18
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 21
cttaggtacg tcaaccat 18

<210> 22
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 22
agcatctaca acatcgac 18

<210> 23
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 23
ggccacgcgt cgactagtac 20

<210> 24
<211> 36
<212> DNA
<213> Artificial

<220>
<223> Primer

<220>
<221> misc_feature
<222> (24)..(25)
<223> N in position 24-25 means deoxyinosine

<220>
<221> misc_feature
<222> (29)..(30)
<223> N in position 29-30 means deoxyinosine

<220>
<221> misc_feature
<222> (34)..(35)
<223> N in position 34-35 means deoxyinosine

<400> 24
ggccacgcgt cgactagtac gggnngggnn gggnng 36

<210> 25
<211> 1596
<212> DNA
<213> Nocardiopsis sp. NRRL 18262

<220>
<221> CDS
<222> (900)..(1466)

<400> 25
acgtttggta cgggtaccgg tgtccgcatg tggccagaat gccccttgc gacagggaac 60

ggattcggtc ggtagcgcat cgactccgac aaccgcgagg tggccgttcg cgtcgccacg 120

ttctgcgacc gtcatgcgac ccatcatcgg gtgaccccac cgagctctga atggtccacc 180

gttctgacgg tctttccctc accaaaacgt gcacctatgg ttaggacgtt gtttaccgaa 240

tgtctcggtg aacgacaggg gccggacggt attcggcccc gatcccccgt tgatcccccc 300

aggagagtag ggaccccatg cgaccctccc ccgttgtctc cgccatcggt acgggagcgc 360

tggccttcgg tctggcgctg tccggtaccc cgggtgccct cgcggccacc ggagcgctcc 420

cccagtcacc caccccggag gccgacgcgg tctccatgca ggaggcgctc cagcgcgacc 480

tcgacctgac ctccgccgag gccgaggagc tgctggccgc ccaggacacc gccttcgagg 540

tcgacgaggc cgcggccgag gccgccgggg acgcctacgg cggctccgtc ttcgacaccg 600

agagcctgga actgaccgtc ctggtcaccg atgccgccgc ggtcgaggcc gtggaggcca 660

ccggcgccgg gaccgagctg gtctcctacg gcatcgacgg tctcgacgag atcgtccagg 720

agctcaacgc cgccgacgcc gttcccggtg tggtcggctg gtacccggac gtggcgggtg 780

acaccgtcgt cctggaggtc ctggagggtt ccggagccga cgtcagcggc ctgctcgcgg 840

acgccggcgt ggacgcctcg gccgtcgagg tgaccacgag cgaccagccc gagctctac 899

gcc gac atc atc ggt ggt ctg gcc tac acc atg ggc ggc cgc tgt tcg 947
Ala Asp Ile Ile Gly Gly Leu Ala Tyr Thr Met Gly Gly Arg Cys Ser
1 5 10 15

gtc ggc ttc gcg gcc acc aac gcc gcc ggt cag ccc ggg ttc gtc acc 995
Val Gly Phe Ala Ala Thr Asn Ala Ala Gly Gln Pro Gly Phe Val Thr
20 25 30

gcc ggt cac tgc ggc cgc gtg ggc acc cag gtg acc atc ggc aac ggc 1043
Ala Gly His Cys Gly Arg Val Gly Thr Gln Val Thr Ile Gly Asn Gly
35 40 45

agg ggc gtc ttc gag cag tcc gtc ttc ccc ggc aac gac gcg gcc ttc 1091
Arg Gly Val Phe Glu Gln Ser Val Phe Pro Gly Asn Asp Ala Ala Phe
50 55 60

gtc cgc ggt acg tcc aac ttc acg ctg acc aac ctg gtc agc cgc tac 1139
Val Arg Gly Thr Ser Asn Phe Thr Leu Thr Asn Leu Val Ser Arg Tyr
65 70 75 80

aac acc ggc ggg tac gcc acg gtc gcc ggt cac aac cag gcc ccc atc 1187
Asn Thr Gly Gly Tyr Ala Thr Val Ala Gly His Asn Gln Ala Pro Ile
85 90 95

ggc tcc tcc gtc tgc cgc tcc ggc tcc acc acc ggt tgg cac tgc ggc 1235
Gly Ser Ser Val Cys Arg Ser Gly Ser Thr Thr Gly Trp His Cys Gly
100 105 110

acc atc cag gcc cgc ggc cag tcg gtg agc tac ccc gag ggc acc gtc 1283
Thr Ile Gln Ala Arg Gly Gln Ser Val Ser Tyr Pro Glu Gly Thr Val
115 120 125

```
acc aac atg acc cgg acc acc gtg tgc gcc gag ccc ggc gac tcc ggc      1331
Thr Asn Met Thr Arg Thr Thr Val Cys Ala Glu Pro Gly Asp Ser Gly
    130                 135                 140

ggc tcc tac atc tcc ggc acc cag gcc cag ggc gtg acc tcc ggc ggc      1379
Gly Ser Tyr Ile Ser Gly Thr Gln Ala Gln Gly Val Thr Ser Gly Gly
145                 150                 155                 160

tcc ggc aac tgc cgc acc ggc ggg acc acc ttc tac cag gag gtc acc      1427
Ser Gly Asn Cys Arg Thr Gly Gly Thr Thr Phe Tyr Gln Glu Val Thr
                165                 170                 175

ccc atg gtg aac tcc tgg ggc gtc cgt ctc cgg acc tga tccccgcggt      1476
Pro Met Val Asn Ser Trp Gly Val Arg Leu Arg Thr
            180                 185

tccaggcgga ccgacggtcg tgacctgagt accaggcgtc cccgccgctt ccagcggcgt   1536

ccgcaccggg gtgggaccgg gcgtggccac ggccccaccc gtgaccggac cgcccggcta   1596
```

<210> 26
<211> 188
<212> PRT
<213> Nocardiopsis sp. NRRL 18262

<400> 26

```
Ala Asp Ile Ile Gly Gly Leu Ala Tyr Thr Met Gly Gly Arg Cys Ser
1               5                   10                  15

Val Gly Phe Ala Ala Thr Asn Ala Ala Gly Gln Pro Gly Phe Val Thr
            20                  25                  30

Ala Gly His Cys Gly Arg Val Gly Thr Gln Val Thr Ile Gly Asn Gly
        35                  40                  45

Arg Gly Val Phe Glu Gln Ser Val Phe Pro Gly Asn Asp Ala Ala Phe
    50                  55                  60

Val Arg Gly Thr Ser Asn Phe Thr Leu Thr Asn Leu Val Ser Arg Tyr
65                  70                  75                  80

Asn Thr Gly Gly Tyr Ala Thr Val Ala Gly His Asn Gln Ala Pro Ile
                85                  90                  95

Gly Ser Ser Val Cys Arg Ser Gly Ser Thr Thr Gly Trp His Cys Gly
            100                 105                 110

Thr Ile Gln Ala Arg Gly Gln Ser Val Ser Tyr Pro Glu Gly Thr Val
        115                 120                 125

Thr Asn Met Thr Arg Thr Thr Val Cys Ala Glu Pro Gly Asp Ser Gly
    130                 135                 140

Gly Ser Tyr Ile Ser Gly Thr Gln Ala Gln Gly Val Thr Ser Gly Gly
145                 150                 155                 160
```

Ser Gly Asn Cys Arg Thr Gly Gly Thr Thr Phe Tyr Gln Glu Val Thr
165 170 175

Pro Met Val Asn Ser Trp Gly Val Arg Leu Arg Thr
180 185